Bayesian Image Restoration and Bacteria Detection in Optical Endomicroscopy

Ahmed Karam Mohammed AbdElkarim

Electrical, Electronic and Computer Engineering Department
School of Engineering and Physical Sciences
Heriot-Watt University
Edinburgh, United Kingdom

Under the supervision of

Prof. Stephen McLaughlin
Dr. Yoann Altmann

A Thesis Submitted for the Degree of
Doctor of Philosophy (PhD) in Electrical Engineering

· September 2018 ·

Abstract

Optical microscopy systems can be used to obtain high-resolution microscopic images of tissue cultures and *ex vivo* tissue samples. This imaging technique can be translated for *in vivo, in situ* applications by using optical fibres and miniature optics. Fibred optical endomicroscopy (OEM) can enable optical biopsy in organs inaccessible by any other imaging systems, and hence can provide rapid and accurate diagnosis in a short time. The raw data the system produce is difficult to interpret as it is modulated by a fibre bundle pattern, producing what is called the "honeycomb effect". Moreover, the data is further degraded due to the fibre core cross coupling problem. On the other hand, there is an unmet clinical need for automatic tools that can help the clinicians to detect fluorescently labelled bacteria in distal lung images.

The aim of this thesis is to develop advanced image processing algorithms that can address the above mentioned problems. First, we provide a statistical model for the fibre core cross coupling problem and the sparse sampling by imaging fibre bundles (honeycomb artefact), which are formulated here as a restoration problem for the first time in the literature. We then introduce a non-linear interpolation method, based on Gaussian processes regression, in order to recover an interpretable scene from the deconvolved data. Second, we develop two bacteria detection algorithms, each of which provides different characteristics. The first approach considers joint formulation to the sparse coding and anomaly detection problems. The anomalies here are considered as candidate bacteria, which are annotated with the help of a trained clinician. Although this approach provides good detection performance and outperforms existing methods in the literature, the user has to carefully tune some crucial model parameters. Hence, we propose a more adaptive approach, for which a Bayesian framework is adopted. This approach not only outperforms the proposed supervised approach and existing methods in the literature but also provides computation time that competes with optimization-based methods.

Acknowledgements

I would like to acknowledge those who have made the production of this tome easier through their help and support. First of all, I would like to express my deepest gratitude to my supervisors, Professor Stephen McLaughlin and Dr. Yoann Altmann, under whose supervision I have been very fortunate to work. Their continuous support and enlightened guidance on both personal and professional matters has been greatly appreciated, and instrumental in helping me to make the most of the past three years. I am also indebted to Heriot-Watt University in Edinburgh for providing me with a James-Watt Scholarship in order to undertake this PhD programme. Furthermore, I would also like to thank Professor Tom Vercauteren and Dr. João Mota for agreeing to be my external and internal examiners.

I am grateful to have been a part of the EPSRC-Proteus multidisciplinary project, all of the members of which I would like to thank. In particular, Professor Mark Bradley, the director of the project, Professor Kevin Dhaliwal, the Proteus Hub team leader, Dr. Anne Moore, the manager of the project, and the rest of the team, who are too numerous to list here, without whom this PhD programme would not have existed.

Dr. Antonios Perperidis, Dr. Paul McCool, Dr. Nikola Krstajic and Dr. Abderrahim Halimi, for numerous enlightening discussions. I would also like to thank Dr. Ahsan Akram and Professor Kevin Dhaliwal for providing the clinical datasets I required in my work, and for the many days they spent annotating the bacteria detection datasets. Thank you also to my office colleagues, past and present, for all the many discussions about our work.

Last, but not least, I owe a debt of gratitude to my mother, late father and my siblings Mohammed, Mahmoud and Amr, for all their help, support, patience and encouragement over the years. It was you who made it possible.

ACADEMIC REGISTRY
Research Thesis Submission

Name:	Ahmed Karam Mohammed AbdElkarim		
School:	Engineering and Physical Sciences		
Version: (i.e. First, Resubmission, Final)	Final	Degree Sought:	PhD in Electrical Engineering

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.

Signature of Candidate:		Date:	

Submission

Submitted By (name in capitals):	AHMED KARAM MOHAMMED ABDELKARIM
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by (name in capitals):			
Method of Submission (Handed in to SSC; posted through internal/external mail):			
E-thesis Submitted (**mandatory for final theses)**			
Signature:		Date:	

Please note this form should be bound into the submitted thesis.

Publications Related to the PhD Thesis

International Journal Papers

- [**Eldaly et al. 2018b**] **Ahmed Karam Eldaly**, Y. Altmann, A. Akram, P. McCool, A. Perperidis, K. Dhaliwal and S. McLaughlin, "Bayesian Bacterial Detection", Submitted to *Elsevier Medical Image Analysis*.

- [**Eldaly et al. 2018a**] **Ahmed Karam Eldaly**, Y. Altmann, A. Perperidis, N. Krstajic, T. R. Choudhary, K. Dhaliwal and S. McLaughlin, "Deconvolution and Restoration of Optical Endomicroscopy Images," *IEEE Trans. Comput. Imag.*, vol. 4, no. 2, pp. 194-205, March 2018.

- [**Perperidis et al. 2017**] A. Perperidis, H. E. Parker, **Ahmed Karam Eldaly**, Y. Altmann, K. Dhaliwal, R. R. Thomson, M. G. Tanner, and S. McLaughlin, "Characterization and modelling of inter-core coupling in coherent fiber bundles," *Opt. Express*, vol. 25, no. 10, pp. 11 932–11 953, May 2017.

International Conference Papers

- [**Eldaly et al. 2018d**] **Ahmed Karam Eldaly**, Y. Altmann, A. Akram, A. Perperidis, K. Dhaliwal and S. McLaughlin, "Patch-Based Sparse representation for Bacterial Detection,", Submitted to *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2019.

- [**Eldaly et al. 2018c**] **Ahmed Karam Eldaly**, Y. Altmann, A. Perperidis and S. McLaughlin, "Deconvolution of Irregularly Subsampled Images," in *IEEE Statistical Signal Processing Workshop (SSP)*, Freiburg, Germany, June 2018, pp. 303–307.

Acronyms and Notations

Acronyms

OEM	Optical endomicroscopy.
PET	Positron emission tomography.
SPECT	Single photon emission computed tomography.
CT	Computed tomography.
MRI	Magnetic resonance imaging.
LED	Light emitting diode.
CCD	Charge-coupled device.
FDA	Food and drug administration.
FOV	Field of view.
PSF	Point spread Function.
FFT	Fast Fourier transform.
DCT	Discrete cosine transform.
DWT	Discrete wavelet transform.
MOD	Method of optimal directions.
K-SVD	K-singular value decomposition.
LS	Least squares.
ML	Maximum likelihood.
MLE	Maximum likelihood estimator.
PDF	Probability density function.
RE	Reconstruction error.
RMSE	Root mean square error.
nD	n-dimensional.
STD	Standard-deviation.
i.i.d.	Independent and identically distributed.

TP	True positive.
FP	False positive.
FN	False negative.
TN	True negative.
TPR	True positive rate.
FPR	False positive rate.
ROC	Receiver operating characteristics.
AUC	Area under curve.
PPV	Positive predictive value.
MSSA	*Methicillin-sensitive Staphylococcus aureus.*
OD	Optical density.
BP	Basis pursuit.
MP	Matching pursuit.
OMP	Orthogonal matching pursuit.
SMV	Single measurement vector.
MH	Metropolis-Hastings.
GS	Gibbs sampler.
PCGS	Partially collapsed Gibbs sampler.
GP	Gaussian processes.
ADMM	Alternating direction method of multipliers.
MCMC	Markov chain Monte Carlo.
VB	Variational Bayes.
KL	Kullback-Leibler.
MAP	Maximum *a posteriori*.
MMSE	Minimum mean square error.
LIP	Linear inverse problem.
(G)MRF	(Gaussian) Markov random fields.
det	Determinant.

Notations

Standard notations

y	Scalar value.

$\mathbf{y}$	Column vector.
$\mathbf{Y}$	Matrix.
$(\cdot)^T$	Transpose operator.
$\text{tr}(\mathbf{Y})$	Trace of the matrix $\mathbf{Y}$.
$\mathbf{1_d}$	$d \times 1$ ones vector.
$\mathbf{0_d}$	$d \times 1$ zeros vector.
$\mathbf{I}_d$	$d \times d$ identity matrix.
$\propto$	Proportional to.
$\sim$	Distribution according to.
$\ll$	Much lower.
$\gg$	Much greater.
$\odot$	Hadamard product (element wise multiplication).
$\|\cdot\|_0$	Conventional ℓ_0 norm.
$\|\cdot\|_1$	Conventional ℓ_1 norm.
$\|\cdot\|_2$	Conventional ℓ_2 norm.
$\|\cdot\|_F$	Conventional Frobenius norm.
$\|\cdot\|$	Absolute value.
$\delta(\cdot)$	Delta function.
$\Gamma(\cdot)$	Gamma function.
$\mathcal{G}(\cdot)$	Gamma distribution.
$\mathcal{IG}(\cdot)$	Inverse Gamma distribution.
$\mathcal{B}e(\cdot)$	Beta distribution.
$\mathcal{N}(\cdot)$	Normal distribution.
$\mathcal{N}_{\mathbb{R}^+}(\cdot)$	Truncated normal distribution on $\mathbb{R}^+$.
$\mathcal{I}_{\mathbb{R}^+}$	Indicator function on $\mathbb{R}^+$.
$\exp(\cdot)$	The exponential function.
$\log(\cdot)$	The logarithm function.
$E(\cdot)$	Expectation operator.
$\mathcal{L}(\cdot)$	Augmented Lagrangian.
$p(\cdot)$	Probability density function.
$\hat{(\cdot)}$	Estimated quantity of $(\cdot)$.

Modelling notations

$\mathbf{g}$	Observation vector.
$\mathbf{x}/\mathbf{X}$	Original intensities vector/matrix.
$\mathbf{y}/\mathbf{Y}$	Approximated observation vector/matrix.
$\mathbf{w}/\mathbf{W}$	Noise vector corrupting the observation vector/matrix..
$\mathbf{r}/\mathbf{R}$	Outlier vector/matrix.
$\mathbf{t}/\mathbf{T}$	Outlier amplitude vector/matrix.
$\mathbf{z}/\mathbf{Z}$	Binary labels vector/matrix.
$\boldsymbol{\psi}/\boldsymbol{\Psi}$	Sparse vector/matrix.
$\mathbf{H}$	Convolution matrix.
$\mathbf{C}$	Fibre core spatial blurring convolution matrix.
$\boldsymbol{\theta}$	Unknown parameter vector.
$\boldsymbol{\phi}$	Unknown hyperparameter vector.
$\mathbf{D}$	Dictionary matrix.
$\boldsymbol{\Delta}$	Covariance matrix.
$\mathbf{B}$	Endmember matrix.
$\mathbf{a}$	Abundance vector.
$\boldsymbol{\mu}$	Estimated mean vector of a multivariate Gaussian distribution.
$\boldsymbol{\Sigma}$	Estimated covariance matrix of a multivariate Gaussian distribution.
$s^2_{\mathbf{t}}$	Outliers variance.
$\mu_{\mathbf{t}}$	Outliers mean.
γ^2	MRF Regularization parameter.
$\omega_{\mathbf{t}}$	Outliers presence probability.
$\boldsymbol{\theta}_{\setminus \mathbf{u}}$	Parameter vector $\boldsymbol{\theta}$ whose $\mathbf{u}$ parameter is omitted.
$\sigma^2_{(\cdot)}$	The variance corresponding to the parameter $(\cdot)$.
$\mathbf{D}_{\mathrm{KL}}$	Kullback-Leibler (KL) divergence.
N	Number of pixels in an OEM image.
N_1	Number of fibre cores in an OEM image.

Sampling/Optimization notations

N_{MC}	Length of Markov chain.
N_{Bi}	Burn-in chain length of the Markov chain.
$\mathbf{x}^{(k)}$	k-th sample/iteration of Markov chain $\left(\mathbf{x}^{(k)}\right)_{k=1,\cdot,N_{\mathrm{MC}}}$/sequence $\left(\mathbf{x}^{(k)}\right)_{k=1,\cdot,N_{\max}}$.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Contents

1.1 Aims and Objectives of the Thesis

A major cause of mortality and morbidity in intensive care patients is ventilator assisted pneumonia (VAP) or the side effects of its treatment [42]. The necessity for prompt information about pneumonia-causing bacteria is a clinically unmet need [89]. As it is generally difficult or dangerous to move intensive care patients, X-Rays or computed tomography (CT) scans cannot usually be performed, and these techniques can only identify low-resolution 'shadows' of lung conditions and so provide limited information about the lung status. The current method of testing existence of VAP is to culture bronchoalveolar lavage fluid (BALF), which can take 48-72 hours to be analyzed and get a result, during which time drugs and antibiotics are given to the patient, some of which can be unsuited to their underlying condition and prove toxic or even fatal [22]. A more tailored treatment in a shorter time-scale would help to mitigate this.

The aim of the EPSRC-Proteus[1] (Edinburgh, UK) multi-disciplinary project is to create a device that can sense and image at a microscopic level inside the living human lung in order to provide information about internal conditions (e.g. pH) and any bacteria that might be present (see a schematic diagram in Figure 1.1). The main sensor in Proteus is based on fibred wide-field optical endomicroscopy (OEM). This and other fibred confocal OEM imaging devices, such as the

[1] http://proteus.ac.uk/

Cellvizio system by Mauna Kea (Paris, France) [16,105,189] provide sparse and irregularly-spaced intensity readings of the scene due to the irregular packing of the fibre cores in the fibre bundle.

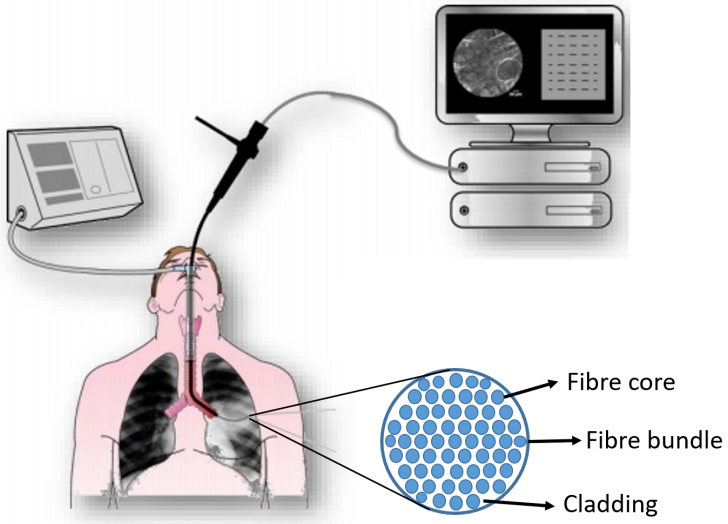


Figure 1.1 Fibred optical endomicroscopy set-up.

This thesis focuses on processing data from fluorescence-based molecular imaging fibred OEM of the distal lung (gas-exchanging alveolar regions). The aim of this PhD project is twofold. Firstly, restoration of the signals at the receiver for better image visualization and post analysis. Secondly, the detection of fluorescently labelled bacteria using both supervised and unsupervised approaches.

Fibre core coupling/talk is a well recognised limitation in coherent fibre bundles, resulting in blurring of the imaged structures. Figure 1.2 provides an illustrative example of cross coupling/talk between fibre cores. If an individual fibre core is illuminated, this results in cross coupling at the neighbouring cores. It can be seen that less coupling occurs at distant cores to the centrally illuminated one and vice versa. Moreover, the irregular packing of the fibre cores in the fibre bundle results in irregularly subsampled grid images which makes it difficult to interpret the imaged scene. Figure 1.3 shows this effect, which is known as the honeycomb effect. The cladding between fibre core causes severe loss of information from the scene under imaging, causing holes in the resulting images.

On the other hand, bacteria can auto-fluoresce when exposed to the appropriate wavelength(s) of light. However the wavelength might not be achievable using the available light source, or might be unsafe to the lung tissue (e.g. ultraviolet light excitation). To address this, bacteria can be stained with fluorophores, or "smartprobe" chemicals [7] which bind to the bacteria to cause fluorescence in response to the light source at the wavelengths at which the device is capable of imaging. Bacteria appear as bright dots in the images and they tend to be as bright as the background structures, making them difficult to identify and quantify.

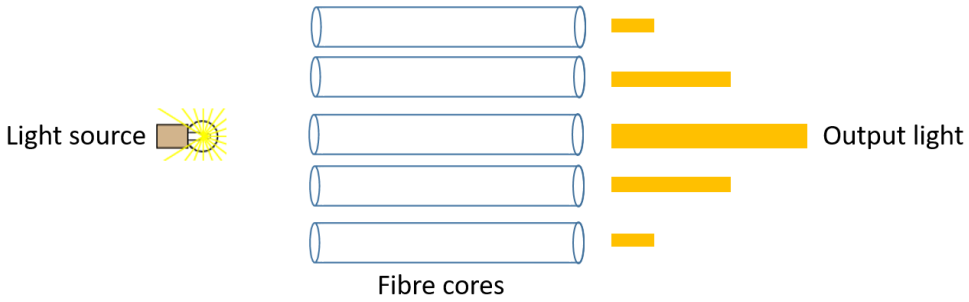


Figure 1.2 A schematic diagram showing the concept of cross coupling/talk between fibre cores. When the middle fibre core is illuminated, this results in cross coupling/talk in neighbouring cores. The amount of coupled light decreases as distance to the central illuminated core increases, and vice versa.

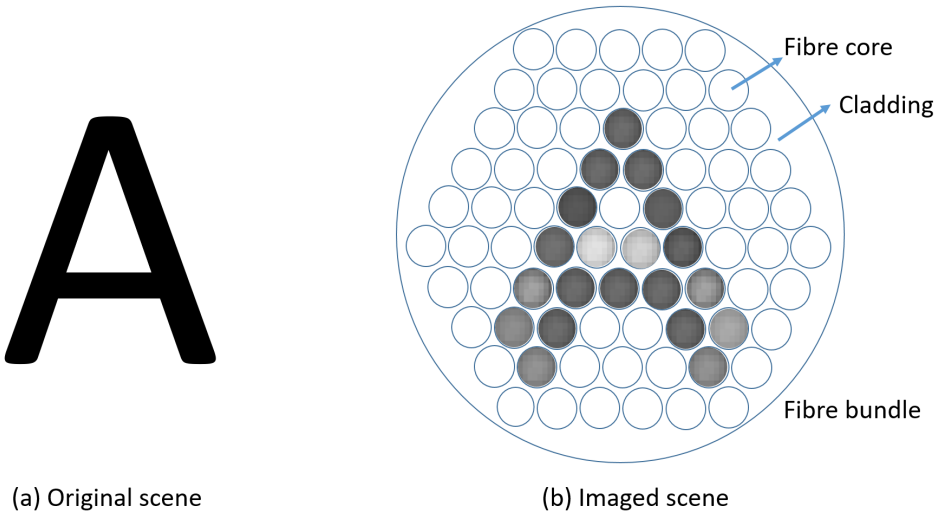


Figure 1.3 A schematic diagram showing the artefacts of the imaged scene caused by irregular packing of fibre cores in the fibre bundle (honeycomb effect). (a) original scene, and (b) imaged scene.

1.2 Main Contributions of the Thesis

The main contributions of this thesis can be seen from two different perspectives: that of the application and that of the methodology.

1.2.1 Contributions from the application perspective

Regarding the application perspective, the contributions can be summarized as follows.

1. We provide a potential solution to the field of view pixelation artefacts caused by the irregular packing of fibre cores in the fibre bundles (honeycomb effect) which limits the lateral and axial resolutions of fibre bundles [**Eldaly et al. 2018a**].

2. We propose a linear model in order to reduce and compensate for the fibre core cross coupling/talk problem, which causes blurring and degradation in the resulting images, using a

deconvolution and restoration mitigation [**Eldaly et al. 2018a**, **Eldaly et al. 2018c**]. To the best of our knowledge, it is the first time to be addressed in this way in the literature.

3. We develop algorithms that can assist clinicians to identify fluorescently labelled bacteria which exist in non-homogeneous backgrounds such as elastin and collagen. As these structures are as bright as bacteria, it is difficult to distinguish them from one another [**Eldaly et al. 2018b**, **Eldaly et al. 2018d**].

1.2.2 Contributions from the methodology perspective

These can be summarized as follows.

1. We propose a new Bayesian model for the deconvolution and restoration of irregularly spatially subsampled images [**Eldaly et al. 2018a**, **Eldaly et al. 2018c**]. The model proposed can be used for different imaging systems as it does not depend on the spatial distribution of the irregularly subsampled intensities. This model assigns suitable prior distributions to the unknown parameters and compares three different estimation strategies, including simulation-based and optimization-based methods. The simulation-based method is fully automatic in the sense that it can estimate all model parameters and hyperparameters, hence the user does not need to set any crucial parameters. However, it exhibits high computational complexity. The optimization-based methods overcome this limitation and provide faster estimations, with one still fully automatic as with the simulation-based method. We validate the model on extensive simulations conducted on both synthetic and real datasets.

2. We develop a non-linear interpolation-based algorithm which is based on Gaussian processes regression to interpolate the deconvolved samples and hence recover a full field of view from the pixeled imaged scene. The algorithm also provides the corresponding confidence intervals of the interpolated pixels [**Eldaly et al. 2018a**].

3. We provide and assess a new parametric blur model for fibre core cross coupling/talk in coherent fibre bundles, based on the spread of light over neighbouring fibre cores [**Eldaly et al. 2018a, Eldaly et al. 2018c, Perperidis et al. 2017**].

4. We develop two new outlier detection (or spot-detection) algorithms [**Eldaly et al. 2018b, Eldaly et al. 2018d**]. The first approach considers the joint formulation of the sparse coding and the outlier detection problems, along with a numerical solver to estimate the unknown model parameters. The approach considers learning background image structures using a dictionary learning-based method and uses this learned model in order to estimate any outlier deviates. The estimation problem is formulated by considering an optimization-based method, where suitable regularization functions are assigned to the unknown model parameters. The proposed model provides fast estimates. However, it requires the user to

set the regularization parameters, which might be difficult to manage should different outlier concentration datasets arise. The second approach considers an appearance model for the outliers and considers a hierarchical Bayesian model in which suitable prior distributions are assigned to the unknown model parameters and their associated hyperparameters. A Markov chain Monte Carlo method based on a partially collapsed Gibbs sampler is then used to sample from the joint posterior distribution. This approach is fully automatic in the sense that the user does not need to set any crucial parameters, hence it is beneficial for datasets where the number of outliers changes from one frame to the other. This algorithm also competes with optimization-based methods as it can provide relatively fast estimates while still being fully automatic.

1.3 Thesis Structure

The rest of the thesis is split into six chapters. The first two chapters provide background information related to fibred OEM, and a review of the literature on existing approaches related to the problems we aim to solve. The next three chapters provide my contributions to image deconvolution and restoration, and spot detection algorithms. One chapter is dedicated to the deconvolution and restoration algorithms and two chapters are dedicated to outlier/spot-detection. Specifically, each chapter is organized as follows.

In Chapter 2, we present the necessary background information related to fibred OEM, which will help us to model the different problems we address in this thesis. The chapter begins by introducing system architecture and the formation of the image using coherent fibre bundles. This allows us to introduce the first problem we tackle in this thesis, which is the pixelation of the resulting images due to the irregular packing of fibre cores (honeycomb effect) that severely limits the lateral resolution of the system. Subsequently, we discuss the two common types of OEM system, which are wide-field and confocal. Then, we introduce the problem of fibre core cross coupling/talk which causes blurring in the resulting images. This problem is hard/expensive to suppress using hardware solutions, hence it is one of the issues we address in this thesis. We aim to recover a non-blurred/non-degraded image from the noisy and blurred one. Moreover, we also discuss the different scanning and imaging modes, the currently available fibred OEM systems being used in clinics and finally the different clinical applications of fibred OEM with emphasis on their ability to identify fluorescently labelled bacteria.

In Chapter 3, we first review the literature on existing methods for image deconvolution and restoration that are used to formulate the fibre core cross coupling/talk problem. Although we focus on Bayesian modelling of the deconvolution and restoration problem, which is the main topic of this thesis, we also provide a quick overview of non-Bayesian methods. This part of the chapter presents how the deconvolution and restoration problem is defined mathematically; outlines existing

approaches to conventional and blind restoration models; surveys the probabilistic models for the observation, the original image, the blur, and their associated unknown parameters, and finally discusses the different inference models used in the literature. The second part of this chapter reviews the literature on most of the existing spot-detection methods that can be adapted to the fibred OEM image formation model, which can be used for the detection of fluorescently labelled bacteria in fibred OEM. These methods can be classified into either unsupervised or supervised. Unsupervised algorithms assume some object appearance model and adjust parameters in order to get the best performance. Supervised methods, on the other hand, learn the object appearance from annotated training data, and then use this learned model in order to provide spot locations on testing data.

In Chapter 4, we provide a statistical model for the fibre core cross coupling problem and the sparse sampling by imaging fibre bundles (honeycomb/pixelation artefact), which are formulated here as a deconvolution and restoration problem as mentioned previously [**Eldaly et al. 2018a, Eldaly et al. 2018c, Perperidis et al. 2017**]. We first provide a complete image model for the fibred wide-field OEM system. However, due to the complexity of this model, we provide an approximation which makes the unknown model parameter inference easier afterwards. We then propose a hierarchical Bayesian model to solve this problem and compare three estimation algorithms to exploit the resulting joint posterior distribution. The first method is based on Markov chain Monte Carlo (MCMC). However, because it exhibits a relatively long computational time, we provide another two optimization-based inference strategies to deal with this issue: in particular, a variational Bayes (VB) approach and an alternating direction method of multipliers (ADMM) algorithm. We then introduce a non-linear interpolation method, based on Gaussian processes regression, in order to interpolate the resulting deconvolved samples to provide a meaningful image. These algorithms are first validated on synthetic datasets to test their performance when ground truth exists and then tested on real wide-field OEM datasets.

In Chapter 5, we propose a joint formulation to the sparse coding and outlier detection problems in OEM images [**Eldaly et al. 2018d**]. The outliers here are considered as fluorescently labelled bacteria, which are immersed in a non-homogeneous background making their detection and differentiation quite difficult. The proposed unsupervised approach considers two main phases: training and testing phases. In the training phase, we learn background image structure appearance (elastin, collagen, etc.) using a dictionary learning step. During the testing phase, we use this learned model in order to predict any outliers deviates. In this approach, we split each image into a set of overlapping patches. Each patch is considered as a linear combination of actual intensity values modelling background structures, which are inferred from the learned dictionary in the training phase, corrupted by an additional term modelling observation noise, and possibly by sparse outlier term modelling bacteria, which are the pixels deviating from the background learned model. We formulate the bacteria detection task with an optimization problem and provide a

numerical solver in order to estimate the unknown model parameters. The algorithm is tested using two real datasets of *ex vivo* lungs instilled with bacteria, in which trained clinicians have identified the locations of bacteria in a set of frames, which is considered to be ground truth.

The bacteria detection approach presented in Chapter 5 although providing good detection performance and outperforming existing methods in the literature, the user has to carefully choose the regularization parameters that control the number of existing outliers. In Chapter 6, we propose a more adaptive approach for the detection of fluorescently labelled bacteria in OEM images [**Eldaly et al. 2018b**]. We consider a Bayesian model which assumes that the observed pixel fluorescence is a linear combination of the actual intensity value associated with elastin, collagen or other background structures, corrupted by additive Gaussian noise and potentially by an additional sparse term modelling bacteria. The bacteria detection problem is formulated in a Bayesian framework and prior distributions are assigned to the unknown model parameters. A Markov chain Monte Carlo method based on a partially collapsed Gibbs sampler is then used to sample the posterior distribution of the unknown parameters. The algorithm is tested using the same two real datasets used in Chapter 5. The proposed model is fully automatic as it does not require the user to set any crucial parameters.

Finally, Chapter 7 provides overall conclusions for this thesis, outlines the limitations of the proposed work and proposes various suggestions for future work.

Chapter 2

Fibred Optical Endomicroscopy

Contents

2.1 Introduction

Clinical molecular imaging techniques such as positron emission tomography (PET), single photon emission computed tomography (SPECT), and fluorescence imaging aim to improve disease diagnosis by using specific SmartProbes for diseases under diagnosis. Due to the advances in imaging techniques, both PET and CT are having an increasing role in diagnosis of different diseases. For example, CT scans can provide useful information about liver, colorectal (cancer that develops in the colon), and lung cancer, however, they don't provide diagnosis on their own as further confirmation by biopsy analysis is required. Moreover, this cumbersome imaging technique is expensive and has low resolution for cellular organisms. Therefore, confirmation by a biopsy is often required to determine the type of cancer at hand.

Most of the time in lung imaging, bronchoscopy is used to perform biopsies, which involves the use of a bronchoscope for viewing the lung airways. Flexible bronchoscopy deploys white light endoscopy for navigation, the diameter of tips of flexible bronchoscopy can be 4 to 6 mm. Over

the last decades, endoscopy techniques have been used to look inside the hollow cavity of the lungs to provide diagnosis or surgery in minimally invasive way [187]. However, they have several limitations in detection of small lesions or determining the underlying microscopic pathological features [139]. The use of bronchoscopy for visualization of respiratory system has been limited so far due to the size of existing bronchoscopes in accessing the central bronchi due to their narrowing diameter. This major limitation has prevented clinicians from gaining a better understanding of peripheral lung cancers and diffuse interstitial diseases. To examine areas of the lungs that are not accessible during a bronchoscopy, clinicians may perform a needle biopsy, i.e., a rather invasive biopsy done from the outside through the chest wall.

Optical endomicroscopy (OEM) is an emerging, fibre-based medical imaging tool, which employs a proximal light source, laser scanning or Light Emitting Diode (LED) illumination, linked to a flexible multi-core fibre, performing microscopic fluorescent imaging at its distal end. The diameter of the packaged fibre can be as small as a few hundred micro meters, enabling real-time imaging of tissues that were previously inaccessible through conventional endoscopy. In pulmonary OEM, auto-fluorescence (at 488nm) generated through the abundant elastin and collagen has enabled the exploration of the distal pulmonary tract [181] as well as the assessment of the respiratory bronchioles and alveolar gas exchanging units of the distal lung [14]. OEM has been used clinically in the lung for the detection of lung cancer [70] and has been used to assess the distal lung [182, 183] including the imaging of parenchymal lung diseases [137]. Fibered Confocal Fluorescent Microscopy (FCFM), also referred to as probe-based confocal laser endomicroscopy (pCLE), is currently the most widely used clinical OEM platform. However, there have recently been a number of studies describing novel, flexible and versatile OEM architectures, employing LED illumination sources, capable of imaging at multiple acquisition wavelengths [98].

In this chapter, we provide an overview of the optical endomicroscopy technology. In particular, we discuss wide-field and confocal endomicroscopy architectures. Moreover, we present the different imaging and scanning modes of optical endomicroscopy. We also provide examples of currently available optical endomicroscopy systems that are being used in clinics/hospitals and their clinical applications not only for human lungs, but also for different organs imaging.

The rest of this chapter is organized as follows. Section 2.2 presents the fibred OEM system architecture and how the image is formed using fibre bundles. Section 2.3 provides in details the different scanning modes of fibred optical endomicroscopy systems. Section 2.4 discusses the two widely used OEM systems. Then, Section 2.5 provides an overview of the currently available OEM in clinics, and Section 2.6 includes the main clinical applications of OEM. The whole chapter is then summarized and concluded in Section 2.7.

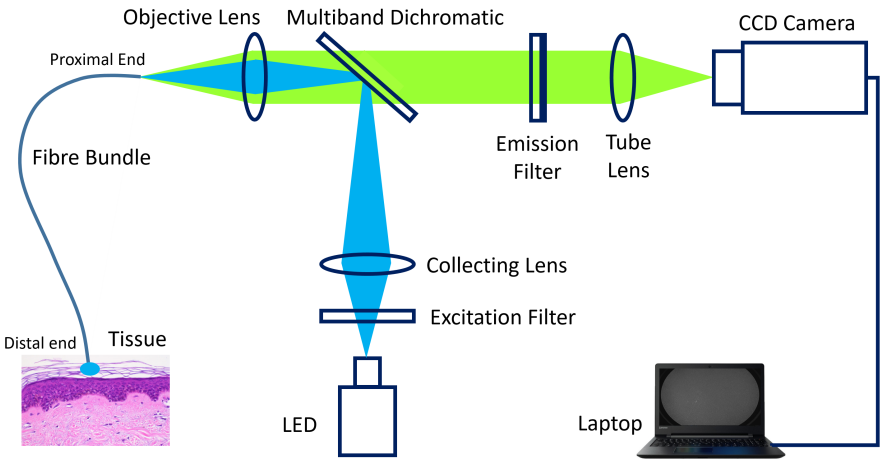


Figure 2.1 Schematic diagram of a single wavelength fluorescence endomicroscope system. Adapted from [98].

2.2 Fibred Optical Endomicroscopy System Architecture

Figure 2.1 depicts a schematic diagram of fluorescence wide-field endomicroscope system. The proximal end of the multicore fibre imaging bundle (closest to the acquisition system) is brought to the focus of the objective enabling fine focusing and optimized coupling of the image from the distal end (closest to the tissue) of the fibre to the rest of the optics. An LED with the required excitation wavelength is used as illumination source coupled to achromatic condenser lenses. The excitation filter, which is a band-pass filter, transmits light with a user-defined wavelength range such that it passes only the wavelengths absorbed by the fluorophore, thus minimizing excitation of other sources of fluorescence. Excitation light is then passed by a dichroic filter, which is an edge filter used at an oblique angle of incidence (typically 45°) to efficiently reflect light in the excitation band and to transmit light in the emission band. Fluorescence signals exiting the objective lens pass through the emission filter before it reaches the camera. The emission filter, which is a band-pass filter, passes only the wavelengths emitted by the fluorophore and blocks all undesired light outside this band. The exiting light is then imaged onto the CCD camera. A camera is used in conjunction with an appropriate band filter set. The camera is then connected to the computer. The system can be extended to multiple wavelength excitations to give multispectral images, rather than single band ones [98].

Excitation light in a miniaturized setting is delivered to the object being imaged through optical fibre bundle. As shown in Figure 2.2, individual optical fibres typically consist of a cylindrical glass core of refractive index n_1, surrounded by a cladding region of slightly lower refractive index n_2 to that of the core (e.g., $n_1 > n_2$). Propagation of light through the fibre is based on the total internal light reflection concept with an incident angle within a certain cone of acceptance at the core-cladding interface, due to the difference in refractive indices between the fibre core and the

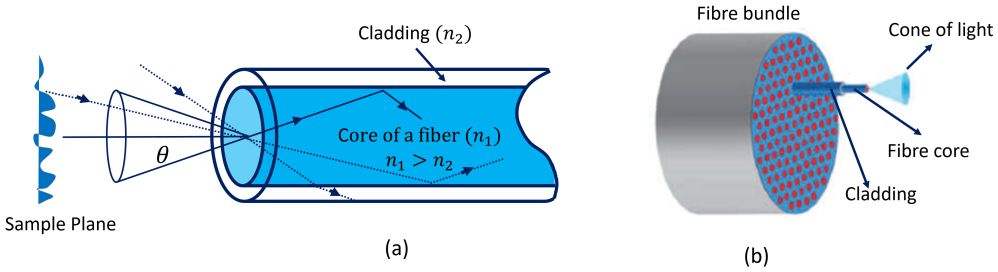


Figure 2.2 Coupling of light to the fibre and optical fibre bundle. (a) Only the light in the core can be transmitted with no leakage; (b) a fibre bundle contains many fibres, ranging from less than 2,000 to as many as 100,000. Figure adapted from [195]

cladding.

Current fibre bundles can contain tens of thousands of $2-4\,\mu m$ diameter fibre cores with core-to-core distance of $3.2-6.0\,\mu m$ and a total diameter of the fibre bundle of hundreds of micrometers to a few millimetres [139]. Coherent fibre bundles are widely used in optical endomicroscopy, where the individual fibre cores preserve their relative spatial relationship at the distal and the proximal ends of the bundle. Optical fibres are compact, lightweight and very flexible and this allow fibre bundles to have the ability to directly transfer images from their distal end to their proximal end in imaging sites where bulk optics can not access. This allows the miniaturization of the imaging probes and to image *in vivo, in situ* biological samples in a non or less invasive way [68, 114]. Moreover, optical fibres are also electromagnetic interference-proof allowing endomicroscopes based on fibre optics to be used in conjunction with CT, MRI, and other imaging modalities. However, the main drawback of conventional fibre bundles is the honeycomb/pixelation artefacts due to the non-imaged spaces between fibre cores, called cladding, which leads to loss of information in the scene to be imaged. As a result, fibre bundles have limited lateral resolution in the sample plane (see figure 2.2). Various image processing algorithms have already been developed in the literature in order to reduce honeycomb/pixelation artefacts [16, 189, 190]. Moreover, we provide another potential solution as we will see in Chapter 4.

Figure 2.3 depicts a background OEM image, which was acquired using the wide-field OEM system [98] being developed by Proteus, which will be discussed in detail in the next section. This image represents a sample presenting constant intensity (constant fluorescence level and uniform excitation), and a zoomed-in region of this image. Bright and dark areas represent fibre cores and the cladding, respectively. This fibre bundle contains cores that are transmitting light and collecting fluorescence light simultaneously. Fibre cores contain information about the object being imaged, while the cladding (the space between the cores) does not. The spread of light in fibre cores causes spatial blurring of imaged objects. The underlying fluorescence profile is usually recovered by interpolating the fibre core intensities [59, 105, 190], i.e., the average or maximum intensities of each core.

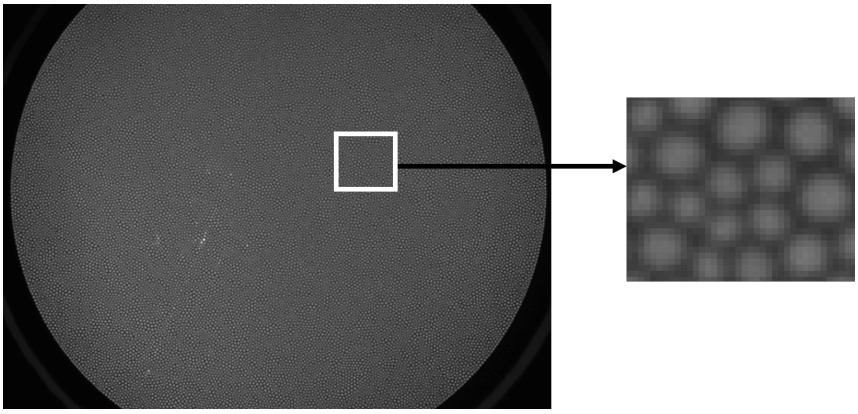


Figure 2.3 Fiber bundle as acquired by the CCD camera with a zoomed inset to the right showing fibre cores (bright) and cladding (dark regions between the cores). This image was acquired at full resolution (1920×1200), using the wide-field OEM system [98].

2.3 Imaging Modes of Fibred Optical Endomicroscopy

Optical endomicroscopy can operate in either fluorescence or reflectance mode. The fluorescence mode is the most commonly used mode in biological imaging, which allows functional imaging of a specific tissue with a good contrast by fluorescence staining. Fluorescence is the process when a photon is absorbed by a suitable molecule, giving up its energy to an electron and thus boosting the electron to a higher energy state. In this context, "suitable" means that there is a possible electronic transition requiring an amount of energy close to that carried by the photon. After a molecule absorbs light, it can return to the ground state with losing the energy it absorbed. Because some of its excitation energy has been lost by this stage, fluorescence occurs at longer wavelength than absorption (see Figure 2.4). Biological molecules can be stained by fluorophores and the fluorescence signal can enable sensitive and quantitative detection of the molecule. There has been an effort to develop molecularly targeted fluorescent probes that can bind and amplify fluorescence in the presence of specific type of tumours [83, 95, 143], inflammation [15], bacteria [7] and fibrogenesis [14], such as peptides [34, 83], antibodies [143] and nanoparticles [26]. This emerging requirement for low limits of detection becomes of paramount importance when imaging small targets such as bacteria superimposed upon highly auto-fluorescent structures like elastin strands in the distal lung [98].

When the optical endomicroscope is used in the fluorescence mode, the light that comes back from the object being imaged is a mixture of reflected and fluorescent signals. To consider only information from a specific tissue, a chromatic filter is usually placed in front of the detector. It is also possible to stain the object being imaged by different fluorophore combination and get multispectral imaging, where each band gives a distinct information about the object. In this thesis, however, we only consider single band images.

On the other hand, reflectance mode has been also used in biological imaging by considering

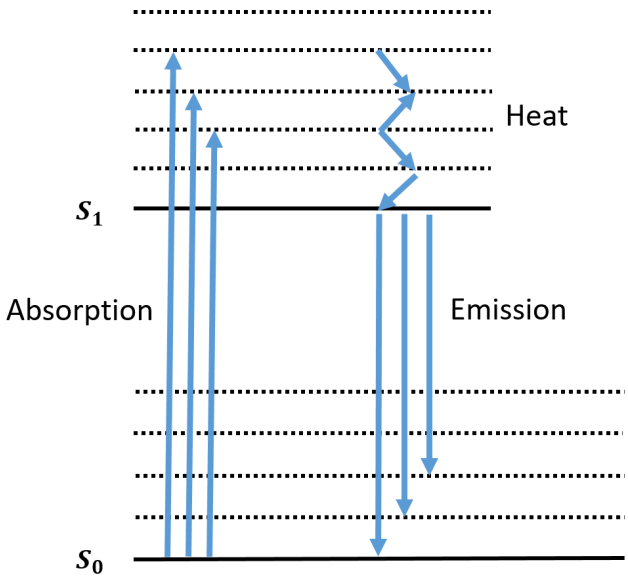


Figure 2.4 A diagram showing fluorescence emission. No fluorescence can be emitted until electrons have reached the lowest vibrational level of S_1.

reflected laser signals rather than fluorescent signals. Optical endomicroscopy can acquire morphological information from tissues without the need of SmartProbes or tissue preparation. The main disadvantage of reflectance-based optical endomicroscopy is its limitation to providing only morphological information about the object being imaged.

2.4 Types of Optical Endomicroscopy Systems

Based on the way individual fibre cores are illuminated, most optical endomicroscopy systems can be classified into either wide-field or confocal endomicroscopes.

2.4.1 Wide-field endomicroscopy

In wide-field endomicroscopy, the entire field of view (FOV) is illuminated simultaneously, and the image is acquired by a 2D array detector, such as a CCD camera. The wide-field method does not provide optical sectioning (3D imaging) since light from different tissue depths, both in and out of focus, contributes to image formation and results in poor contrast (see Figure 2.5). Whether a structure can be visualized depends on how bright it will appear and whether it is strong enough to be visualized above the background intensity. Because no scanning is required, wide-field imaging has the advantage of simpler and less expensive instrumentation and faster full-frame acquisition [98, 197]. The main drawback of this technique is the blurring effect caused by crosstalk between adjacent fibre cores, and off-focus light coming from different depths [59, 67, 146]. Figure 2.6 shows an example of the USAF resolution test chart imaged by the wide-field fibred OEM system in [98], and the corresponding natural neighbour interpolation [166, Chapt. 2] of the

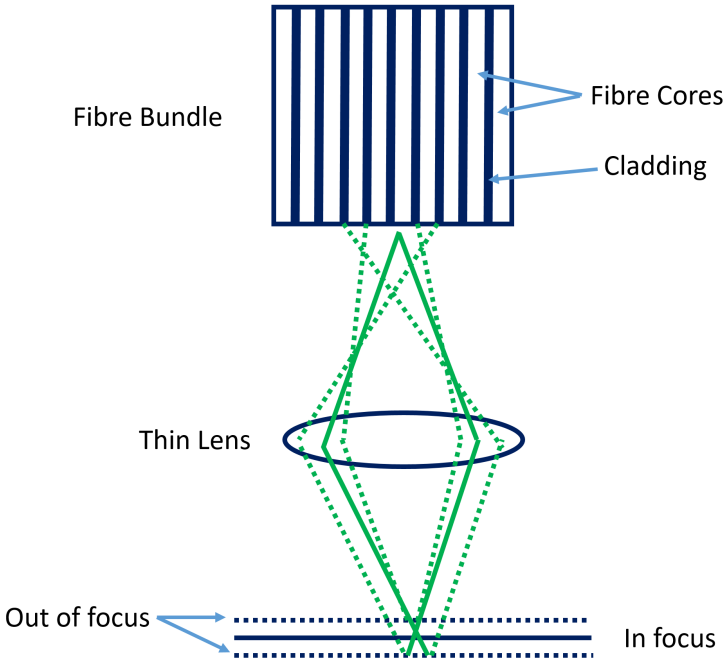


Figure 2.5 In and out of focus fluorescence signals. In wide-field endomicroscopes, the two types of signals contribute to form the final image, while confocal endomicroscopes blocks out of focus signals. This diagram is adapted from [189].

fibre core intensities. We can observe the blurring which is caused by the cross coupling effect as well as the sparsity of the data. In this thesis, as we will see in Chapter 4, we developed several algorithms to compensate for this effect and hence enhance the lateral resolution of the resulting images.

2.4.2 Confocal endomicroscopy

In confocal endomicroscopy, a point in the sample is illuminated by a focused laser spot and the emitted light from this point is detected through a confocal pinhole. Depending on the imaging mode the detector either measures the fluorescence light or the reflected light. The point is then scanned laterally in two dimensions for 2D imaging as well as axially to build up a 3D-image. Conceptually, a straightforward way to convert the fibre bundle endomicroscope into a confocal system is to illuminate the individual fibres in the bundle one at a time and detect the remitted light coming back through the same fibre core while rejecting light coming back through adjacent fibres that carry multiply scattered photons and signals from out-of-focus objects (see Figure 2.5). To obtain a 2D image, different fibres in the fibre bundle can be illuminated sequentially and each fibre detected confocally. In the simplest implementation, a focused laser is scanned continuously over the entrance surface of the fibre bundle in a 2D (x-y) raster pattern. This raster scan pattern is relayed to the distal end of the bundle and imaged onto the sample by the miniature objective length. Although the ability of this system in reducing out-of-focus light, the blurring effect caused

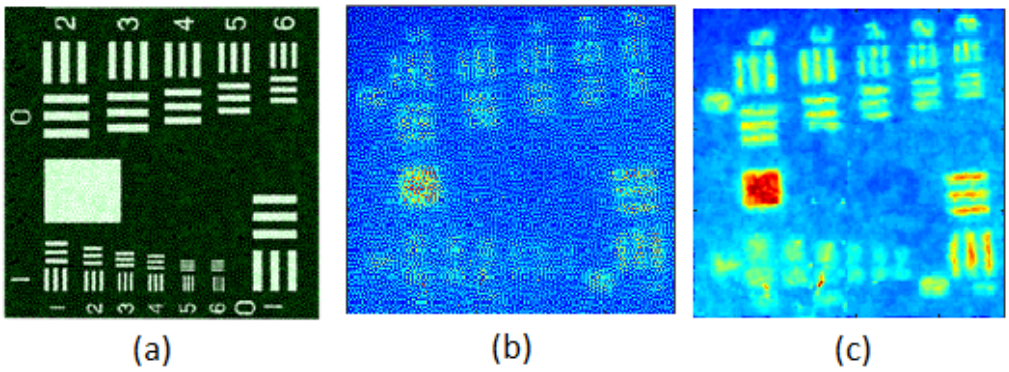


Figure 2.6 Example of cross coupling effect on OEM images of the USAF chart. (a) Binary image of USAF chart, fibre-core locations overlaid, (b) the chart imaged by the wide-field optical endomicroscopy system [98] and (c) natural neighbour interpolation [166, Chapt. 2] of the fibre core intensities of the image in (b).

by fibre core cross coupling still exists due to the internal coupling between fibre cores.

Confocal optical endomicroscopes employ, based on the location of scanner from an endoscopist's point of view, either proximal or distal scanning. In proximal scanning, the imaging probe does not contain scanner and there is no restriction on the size of scanner. In contrast, distal scanning requires developing miniature endoscopic probe that contains scanning unit. In proximal scanning, each individual fibre within a fibre bundle is used as the pinhole while scanning is performed across the fibre bundle. The advantage of this method is the unlimited size of scanning system as it is not included in the endomicroscopic probe. Proximal scanning systems for confocal imaging use one or more moving mirrors to scan the confocal pinholes across the fibre bundle [139].

2.5 Currently Available Fibred Optical Endomicroscopy Systems

The Cellvizio system developed by Mauna Kea Technologies, Paris-France, is the most commonly used commercial confocal endomicroscopy system currently being used in clinics. It is a standard imaging system that is based on a fibre bundle that can be passed through the accessory channel of any standard endoscope, allowing for fast image acquisition *in vivo, in situ* in real time (more than 12 frames per second), along with a wide range of compatible multi-core fibre probes with diameters smaller than 0.3 mm, approximate lateral resolution smaller than $< 3.3\,\mu m$ and field of views between 300 and 600 µm. The Cellvizio system utilizes a proximally scanned fibre bundle to deliver excitation laser light to the sample. It has been used for real-time microscopic imaging of gastrointestinal, urinary tract, and respiratory epithelium, enabling real time imaging of moving structures, making Cellvizio the most widely used endomicroscopy platform approved for clinical use [40, 102, 183, 202]. On the other hand, there is currently a promising optical endomicroscopy

system being developed by Proteus (Edinburgh-UK), which is wide-field fluorescence-based. It can enable multiplexed molecular imaging in the alveolar space of human lung tissue [98].

2.6 Clinical Applications of Fibred Optical Endomicroscopy

Optical endomicroscopy can be used in real-time diagnosis processes of organs that were previously inaccessible by other means, and thus can eliminate the need for tissue biopsies. Reducing the need for histological tissue biopsies is important specially for organs with possibly high morbidity and infection risk as in lymph nodes or for organs with potential to inflict irreparable damage as in the brain. Furthermore, in the cases when histological tissue biopsies are required, optical endomicroscopy can help identify these locations while minimizing trauma. Due to the sufficiently small size of fibre bundle probes, they can access several areas within the body that were previously inaccessible, *in vivo, in situ*, and thus can provide optical biopsies in a wide and ever-increasing range of organ systems, such as the bladder, ovaries, the cervix, the gastrointestinal tract, the oral cavity and the lungs. Usually, the fibre bundle is guided through the working channel of an endoscope to a region of interest and intravenous or SmartProbes are employed to induce and enhance fluorescence of the imaged structure.

Up-to-date, the most widely use of fibre bundle optical endomicroscopy is in providing optical biopsies for the Gastro-Intestinal (GI) tract [71,194,196]. In the lower Gastro-Intestinal tract, fibred optical endomicroscopy has been used to test the activity of inflammatory Bowel disease [151,159]. Moreover, it has been used to detect malignancy in colorectal polyps [50] and detect colonic neoplasia [180]. On the other hand, fibre-endomicroscopy has been employed in the upper GI tract to polyps and neoplastic lesions as well as gastritis and metaplastic lesions in the stomach and duodenum and structural changes in the oesophagus mucosa associated with squamous cell carcinoma and Barrett's oesophagus [55]. In pulmonology, elastin and collagen generate auto-fluorescent signals that have permitted the assessment of the alveolar gas exchange units and the respiratory bronchioles as well as the exploration of the distal pulmonary tract without the need for exogenous contrast agents. Several studies have demonstrated the ability of fibred confocal fluorescence endomicroscopy to detect a wide range of pathologies including acute lunge cellular rejection (AR) following lung transplantation [207], changes in cellularity in the alveolar space as indicator of pneumonia, cross-sectional and level of fluorescence changes in the alveolar structure in emphysema [137, 208], elastic fibre distortion [209] and neoplastic changes in epithelial cells [70, 181, 184] in bronchial mucosa. In urology, optical endomicroscopy has been used to image structural epithelial changes observable in bladder neoplasia [176] as well as upper tract urothelial carcinoma [44]. Other clinical applications of fibred optical endomicroscopy include imaging the oropharyngeal cavity by differentiating between healthy, squamous, and squamous cell carcinoma epithelial tissues [1]. It has been also used in imaging of brain tumours by providing immediate

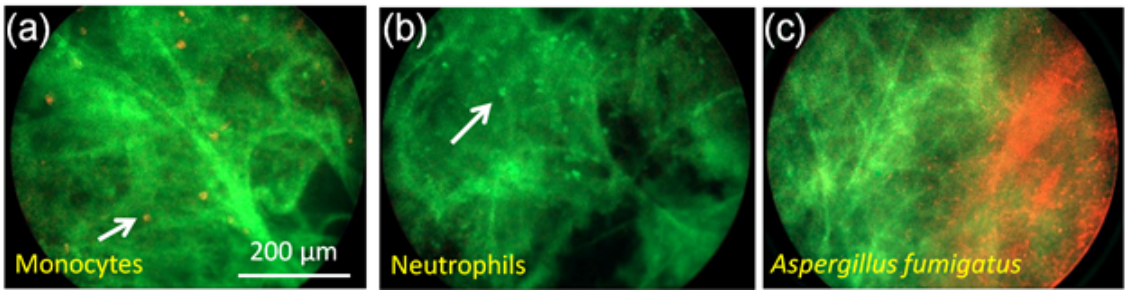


Figure 2.7 Images of multiple targets in ex vivo lung tissue including (a) labelled monocytes, (b) labelled neutrophils and (c) fungus (Aspergillus fumigatus) [98].

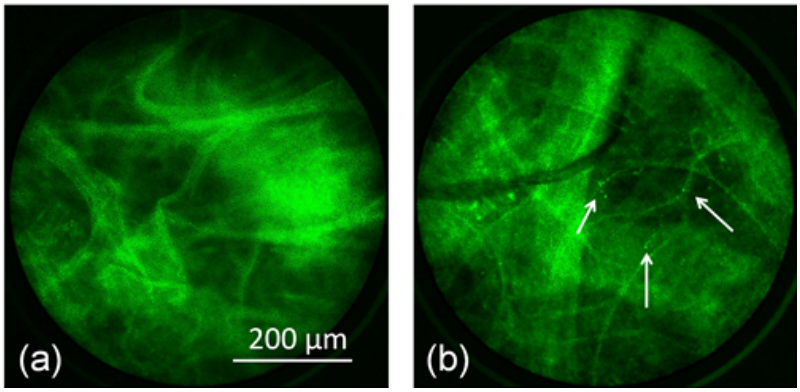


Figure 2.8 Images of ex vivo lung tissue (a) without bacteria (b) and with PKH green labelled bacteria [98].

histological assessment of the brain-to-neoplasm interface and hence improving tumour resection [132,212]. In [92,173], the authors showed that fibred optical endomicroscopy can be used to detect potential malignancy by imaging pancreatobiliary structure.

On the other hand, wide-field fibred optical endomicroscopy has been used in a preliminary study to target inflammation and infection of pulmonary *ex vivo* lung tissues [98]. Figure 2.7 shows the ability of the system to target labelled monocytes, neutrophils, and fungus (e.g., Aspergillus fumigatus). Moreover, Fig. 2.8 shows the ability of the system to view fluorescently labelled bacteria. The right (resp. left) hand image shows an image of *ex vivo* lungs without (resp. with) bacteria. While bacteria can auto-fluoresce when exposed to light at an appropriate wavelength, the emitted fluorescent signal is usually too weak to be imaged effectively. To address this, bacteria can be stained by SmartProbes [7], which make bacteria fluorescence in response to a light excitation at pre-determined wavelengths. Stained bacteria appear as bright dots in the images, whereas elastin and collagen are naturally auto-florescent and display a mesh like structure in the distal lung.

2.7 Summary

In this chapter, we provided a brief overview of fibred optical endomicroscopy, the imaging technique whose data we process in this thesis. First, we provided system architecture and how the image is formed by using fibre bundles, which in turn produces pixellized images due to the irregular packing of the fibre cores that severely limit the lateral resolution of the system. Then, based on fibre bundle illumination, we classified fibred OEM systems into either wide-field-based or confocal-based. Inter-core coupling is a well recognised limitation in coherent fibre bundles in fibred OEM, resulting in blurring of the imaged structures and consequently a worsening in the associated limits of detection. The fibre core cross coupling/talk is hard/expensive to suppress using hardware solutions, hence it is one of the issues we aim to solve in this thesis along with the field of view pixelation effect. Moreover, we discussed the different scanning and imaging modes, the currently available fibred OEM systems and finally its different clinical applications.

To sum up, the main issues with the OEM system we aim to solve are

1. The field of view pixelation artefacts caused by the irregular packing of the fibre cores in the fibre bundles, known as the honeycomb effect. The cladding between fibre core causes sever loss of information from the scene under imaging, which causes holes in the resulting images.

2. The fibre core cross coupling problem which causes blurring and degradation in the resulting images.

3. Detection of fluorescently labelled bacteria which exist in a non-homogeneous background like elastin and collagen, which are as bright as bacteria, making them quite difficult to distinguish.

In Chapter 4, we formulate the fibre core cross coupling and the field of view pixelation problems as a deconvolution and restoration task, which it is the first time to be addressed in this way in the literature. Moreover, in Chapters 5 and 6 we introduce novel spot detection algorithms in order detect fluorescently labelled bacteria in lung images acquired by a confocal OEM system. Hence, in the next chapter, we review the literature on existing methods for image deconvolution, restoration and spot detection.

Chapter 3

Deconvolution, Restoration and Spot Detection: Theory and Applications

Contents

3.1 Overview

The fibre core cross coupling/talk problem mentioned in Chapter 2 has been studied in the literature both experimentally [45, 146, 201] and within a theoretical framework (coupled mode theory) [152, 174], providing recommendations for optimal fibre bundle designs. However, cross coupling can be mitigated yet not eliminated through optimal fibre design. In this thesis, in order to compensate for this effect, we formulated the cross coupling mitigation and the field of view pixelation problems as a deconvolution and restoration task. To the best of our knowledge, it is the first time these problems are addressed in this way. In the first part of this chapter, we review the literature on existing methods for image deconvolution and restoration. These two problems have been addressed in the literature via different methods. Here we focus on Bayesian models, but we also provide a brief overview on non-Bayesian methods. On the other hand, we aim to develop algorithms for automatic detection of fluorescently labelled bacteria. Due to the lack of methods specially those applied to fibred optical endomicroscopy, in the second part of this chapter, we review the literature on most of existing spot detection methods that can be adapted to this imaging technique.

3.2 Image Deconvolution and Restoration

3.2.1 Introduction

In deconvolution and restoration problems, one wants to recover an image from noisy and degraded data; the target image is the ideal representation of the observed scene. However, during the observation process, there may exist uncertainty/deformation in the measurements occurring as blur, noise, etc.. The key idea to solve this ill-posed inverse problem (it might not have a solution

in the strict sense and/or solutions might not be unique) is the introduction of prior knowledge about the original image into the restoration process. Most of the deconvolution and restoration problems introduced in the literature differ mainly in the prior knowledge about the image they intend to restore, and the estimation strategy adopted.

In the literature, two categories of Bayesian restoration methods are found [28]. The first consists of computing an estimate of the original image assuming the blur is known [2–4,64,65,107,200]. The second is a class of image restoration known as blind image restoration, which assumes that both the true image and the blur are unknown [17,30,39,88,101,126,140,158,177,206,210]. In general, the degradation can be non-linear and spatially varying (non uniform motion, imperfect optics); however, in most of the work introduced in the literature, it is assumed that the observed image is the output of a linear spatially invariant (LSI) system to which noise is added.

The rest of this section is organized as follow. In Section 3.2.2 we define mathematically the deconvolution problem. In Section 3.2.3 we provide a classification of the existing approaches to supervised deconvolution. In Section 3.2.4 we formulate the deconvolution problem within the Bayesian framework, and survey the probabilistic models for the observation, the original image, the blur, and their associated unknown parameters. In Section 3.2.5 we discuss solutions to the deconvolution problem as inference models under the Bayesian framework. Finally, in Section 3.2.6 we briefly review deconvolution models which appeared in the literature and cannot be easily obtained using a strict Bayesian formulation.

3.2.2 Problem Formulation of the Deconvolution Problem

The deconvolution problem for digital image processing can be formulated by the following discrete model

$$g(n) = \sum_{s \in S_h} h(n,s)x(s) + w(n), \quad n = (n_1, n_2) \in S_x, \tag{3.1}$$

where $g(n), h(n,s), x(n)$, and $w(n)$ represent the degraded image, the blur function or point spread function (PSF), the original image to be recovered, and the observation noise, respectively, $S_x \in \mathbb{R}^2$ and $S_h \in \mathbb{R}^2$ are the supports of the image and the PSF respectively. The additive noise $w(n)$ is assumed to be white Gaussian independent and identically distributed (iid) noise, which is assumed to be uncorrelated with the image in all of the revised methods. The noise might originate during acquisition, transmission or preprocessing.

The main aim of the supervised deconvolution problem is to estimate the original image $\mathbf{x}$, other deconvolution problems consider estimating both the original image $\mathbf{x}$ as well as the blur function $\mathbf{h}$. These methods are known as blind deconvolution problems. The LSI degradation model can be written as

$$g(n) = (x * h)(n) + w(n) = \sum_{s \in S_h} h(n-s)x(s) + w(n), \tag{3.2}$$

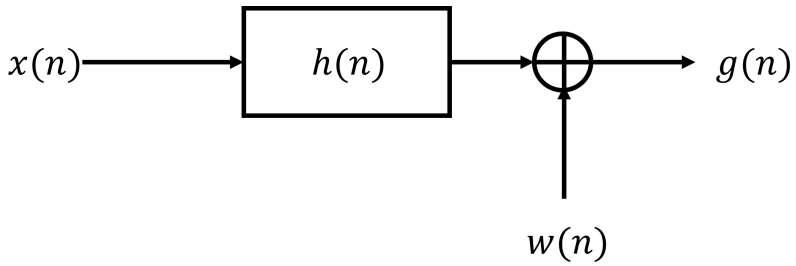


Figure 3.1 A schematic diagram of linear image degradation model.

where $(*)$ denotes the 2-D convolution operator. The block diagram representing this linear degradation model is depicted in Fig. 3.1.

The linear image degradation model represented by Eq.(3.1) and Eq.(3.2) can be written in a matrix form as

$$\mathbf{g} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{3.3}$$

where the vectors $\mathbf{g} \in \mathbb{R}^N$, $\mathbf{x} \in \mathbb{R}^M$, and $\mathbf{w} \in \mathbb{R}^N$ are the observed image, the original image and the observation noise respectively which are all vectorized to form vectors. $\mathbf{H} \in \mathbb{R}^{N \times M}$ represents the blur matrix.

The main objective of the deconvolution problem is to find the estimate $\hat{\mathbf{x}}$ for $\mathbf{x}$ based on $\mathbf{g}$. This is an ill-posed linear inverse problem, as the matrix $\mathbf{H}$ is singular and/or ill-conditioned ($N \leq M$, such that $\mathbf{g} \approx \mathbf{H}\hat{\mathbf{x}}$) hence regularization for the unknown variables must be considered (for $\mathbf{x}$ and $\mathbf{w}$). In blind deconvolution problems, the estimate $\hat{\mathbf{h}}$ is also found for $\mathbf{h}$ by incorporating prior knowledge about $\mathbf{h}$.

3.2.3 Classification of Image Deconvolution Algorithms

Deconvolution problems may be classified based on the way the blur function is identified into either *a priori* known information or jointly estimated within the original image.

3.2.3.1 *A priori*-identified blur methods deconvolution

In these methods, the blur can either be estimated separately from the degraded image, and later used in the estimation of the original image process, or it can be identified by experimental approaches. For the first method, parametric blur models are primarily used as we will see in Section 3.2.4.2, the aim is then to identify the blur parameters from the observed degraded image [29, 37, 179].

Experimental approaches, which are widely used in microscopy, remote sensing, medical ultrasound, etc., are used for estimating the blur function [146]. Images of one or more point sources (Dirac function excitation) are used to provide an estimate of the blur function. This can also be

used as a prior knowledge for estimating jointly the blur as well the original image as we will see in the next subsection.

3.2.3.2 Blind deconvolution methods

In this class, the blur function and the original image are jointly estimated. Prior knowledge about the blur function and the image to be recovered are introduced in the form of statistical models (like the Bayesian models introduced in the next subsection), or penalization/constraints in the cost function to be optimized. Parameters representing such models are also estimated within the inference process [21, 30, 39, 88, 110, 126, 130, 210].

The following subsection illustrates how the deconvolution problem is formulated within a hierarchical Bayesian model in the literature. Most of these methods rely on selecting prior distributions to the unknown model parameters and hyperparameters. The joint posterior distribution is then computed using Bayes' theorem, and then tackled using one of the inference algorithms. The main differences between these methods are the choice of the prior distributions of the unknown model parameters as well as the method used to compute the model estimators.

3.2.4 Bayesian Modelling for Image Deconvolution

Bayesian approaches assume that all model parameters and observed variables are unknown stochastic quantities, by assigning them probability distributions based on prior beliefs. Therefore, in image deconvolution problems, the original image $\mathbf{x}$, and the noise $\mathbf{w}$ (and the blur $\mathbf{h}$ in blind deconvolution) in Eq.(3.3) are all considered as samples of random variables with corresponding probability density functions (PDFs) that model our knowledge about the imaging process and the nature of the images. These distributions may depend on hyperparameters denoted as $\boldsymbol{\phi}$.

Usually, $\boldsymbol{\phi}$ is assumed to be known (estimated separately from $\mathbf{x}$ and $\mathbf{g}$). On the other hand, hierarchical Bayesian models may be adopted where $\boldsymbol{\phi}$ is assumed to be unknown, in which case we also model our prior knowledge of its values. These models offer a flexible and consistent methodology to deal with uncertainty in inference when limited amount of data or prior information is available. This hierarchical modelling allows to write the joint distribution as follows

$$p(\boldsymbol{\phi}, \mathbf{x}, \mathbf{h}, \mathbf{g}) = p(\mathbf{g}|\boldsymbol{\phi}, \mathbf{x}, \mathbf{h}) p(\mathbf{x}, \mathbf{h}|\boldsymbol{\phi}) p(\boldsymbol{\phi}), \tag{3.4}$$

where $p(\mathbf{g}|\boldsymbol{\phi}, \mathbf{x}, \mathbf{h})$ denotes the likelihood of the observations. Typically, the original image to be recovered $\mathbf{x}$ and the blur function $\mathbf{h}$ are assumed to be *a priori*, given the hyperparameter vector $\boldsymbol{\phi}$, independent i.e., $p(\mathbf{x}, \mathbf{h}|\boldsymbol{\phi}) = p(\mathbf{x}|\boldsymbol{\phi}) p(\mathbf{h}|\boldsymbol{\phi})$. The goal is then is to perform Bayesian inference using the full posterior distribution given by

$$p(\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}) p(\mathbf{x}|\boldsymbol{\phi}) p(\mathbf{h}|\boldsymbol{\phi}) p(\boldsymbol{\phi})}{p(\mathbf{g})}. \tag{3.5}$$

Since $p(\mathbf{g})$ is a normalization constant which does not depend on the parameters of interest (e.g., $\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}$), the joint posterior distribution can be written as follows

$$p(\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}) p(\mathbf{x}|\boldsymbol{\phi}) p(\mathbf{h}|\boldsymbol{\phi}) p(\boldsymbol{\phi}), \tag{3.6}$$

where $\propto$ reads "is proportional to". In the following subsection, we define the likelihood function of the observations and provide various prior models for the image, blur, and their corresponding hyperparameters, that have been used in the literature. We then present various Bayesian inference strategies that provide estimators for the unknown model parameters and hyperparameters.

3.2.4.1 Observation model

Following the joint posterior distribution given in Eq.(3.6), the first step of the Bayesian formulation problem is to define the likelihood function of the observed image, $\mathbf{g}$, denoted as $p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi})$. As per Eq.(3.3), the PDF of the observed image given the unknown model parameters is related to that of the observation noise $\mathbf{w}$. A widely used model for $\mathbf{w}$ in the literature is zero mean independent and identically distributed white Gaussian noise, with covariance matrix $\sigma^2 \mathbf{I}_N$, with PDF denoted as $\mathbf{w} \sim \mathcal{N}(\mathbf{w}; \mathbf{0_N}, \boldsymbol{\sigma^2 \mathbf{I}_N})$, where $\sim$ reads "is distributed according to". Consequently, the likelihood function can be written as

$$p(\mathbf{w}) = \quad p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}) = \quad \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left[-\frac{\|\mathbf{g} - \mathbf{H}\mathbf{x}\|_2^2}{2\sigma^2}\right], \tag{3.7}$$

where $\|\cdot\|_2^2$ denotes the conventional squared ℓ_2 norm.

Other noise models, for instance Poissonian noise, which can be used to model low-photon count imaging systems, are also assumed in some deconvolution problems in the literature [64, 117]. We will, however, concentrate on the Gaussian noise model which will be used to model the different problems we considered in this thesis, which has been shown to be well adapted for the data we processed.

3.2.4.2 Parametric prior blur models

Analytical models are frequently used in the literature for defining the degradation operator, i.e, PSF. In this case, the blur's prior distribution is directly parametrized by the parametric model unknowns (the priors are on the blur parameters rather than on $\mathbf{h}$ directly). Often when the parametric model of the PSF is assumed to be known, the unknown parameters of the model can be estimated for example using maximum likelihood (ML) methods [103]. On the contrary, these unknown parameters can also be estimated *a priori* by experimental approaches.

1. *Linear motion blur*: During exposure time, if a camera moves a certain distance, each point of the imaged scene is pictured onto several pixels of the resulting image, which produces a

blurred image along the motion direction. In this case, the blur is due to the incoherence in the relative movement between the camera and the captured scene. In the case of linear homogeneous motion blur, the PSF $h(u, v)$ is a one-dimensional rectangular function [115], which is a line segment through the origin. The angle ϕ of the line segment to the x-axis is equal to the direction of motion, and the length L equals the distance one pixel is moved by the motion. Thus the intensity is $1/L$ along the line and 0 elsewhere. The PSF in this case can be written as

$$h(u,v) = \begin{cases} \frac{1}{L}, & (u,v)\left(\sin(\phi), \cos(\phi)\right)^T = 0, \text{ and } u^2 + v^2 < \frac{L^2}{4}. \\ 0, & \text{otherwise.} \end{cases} \tag{3.8}$$

Figure 3.2(a) shows an example of a discretised motion blur kernel when $\phi = 45$, and $L = 10$.

2. Atmospheric turbulence blur: This type of blur is caused by the random fluctuations of the refraction index of the medium, which lead to blurring in images acquired from a long distance away. It is common in aerial imaging and remote sensing applications. It can be modelled by a Gaussian PSF for long-term exposure through the atmosphere as follows [80]

$$h(u,v) = K \exp\left(-\frac{(u-k)^2 + (v-l)^2}{2\sigma_{\mathbf{h}}^2}\right), \tag{3.9}$$

where (k, l) is the center of the PSF, K is a normalizing constant ensuring that the blur has a unit volume, and $\sigma_{\mathbf{h}}^2$ is the variance that determines the width of the blur. Figure 3.2(b) shows an example of a Gaussian blur kernel when $\sigma_{\mathbf{h}}^2 = 5$.

3. *Out-of-focus blur*: Blurring of images due to out of camera focus primarily arises due to the finite size of the camera aperture. The image formation model produced by the camera depends on many parameters, including the camera aperture size and shape, the distance between object and camera, the focal length, the wavelength of the incoming light, and effects due to diffraction [106]. When the blur due to defocusing is large, the uniform circular PSF function is used as an approximation to these effects

$$h(u,v) = \begin{cases} \frac{1}{\pi r^2}, & (u-k)^2 + (v-l)^2 \leq r^2. \\ 0, & \text{otherwise.} \end{cases} \tag{3.10}$$

where r is the radius of the blur. Figure 3.2(c) shows an example of a circular blur kernel when $r = 4$.

4. *Crosstalk blur in fibred optical endomicroscopy*: Fiber core cross coupling is one of the main reasons for image degradation in optical endomicroscopy [59, 60, 152, 201]. If an individual fibre core is illuminated, the neighbouring cores will be affected by a specific percentage of the

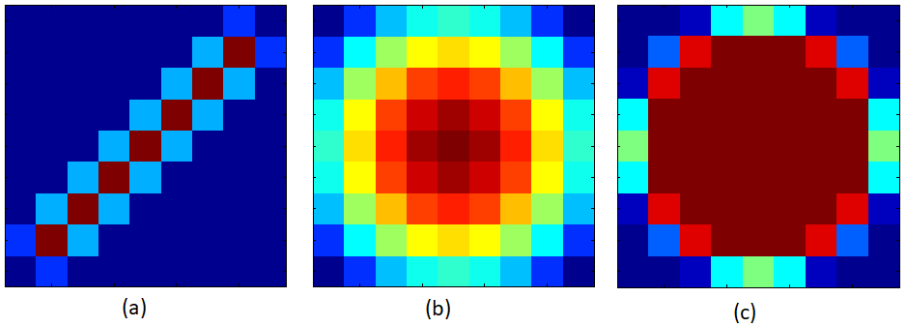


Figure 3.2 Examples of PSF kernels each of size 9×9. (a) motion blur kernel, with $\phi = 45$ and $L = 10$, (b) Gaussian blur kernel, with $\sigma^2_{\mathbf{h}} = 5$, and (c) circular blur kernel, with $r = 4$.

incident light on the illuminated core. There have been several studies that have quantified the average spread of inter-core coupled light, with approximately a third of the overall light coupling to neighbouring cores [146]. In [**Eldaly et al. 2018a, Eldaly et al. 2018c,**], cross-coupling between fibre cores is modelled by an isotropic zero mean 2D generalized Gaussian kernel as follows

$$h(i,j) = \exp\left(-\left(\frac{d_{i,j}}{\alpha_{\mathbf{h}}}\right)^{\beta_{\mathbf{h}}}\right), \tag{3.11}$$

where $d_{i,j}$ denotes the euclidean distance between the cores (or spatial locations) i and j. From Eq.(3.11), it can be seen that neighbouring fibre cores will be more closely coupled than distant ones. The values of $\alpha_{\mathbf{h}}$ and $\beta_{\mathbf{h}}$, which control the amount of cross-coupling (the higher, the more coupling) and which are system dependent can be adjusted from preliminary measurements (calibration). Note however that background offset and gain [161], arising due to the difference in fibre core sizes are neglected here to make the formation of the restoration forward model and the estimation of the unknown model parameters tractable, as we will see in Chapter 4. This parametric prior blur model has been investigated during this thesis and is one of the contributions [**Eldaly et al. 2018a, Eldaly et al. 2018c, Perperidis et al. 2017**].

3.2.4.3 Image and blur prior models

Due to the ill-posed nature of the deconvolution problem, prior beliefs should be assigned to the image to be recovered $\mathbf{x}$ and the blur function $\mathbf{h}$ in case of blind deconvolution problems. These beliefs can be formulated using prior distributions $p(\mathbf{x}|\boldsymbol{\phi})$ and $p(\mathbf{h}|\boldsymbol{\phi})$, respectively. These distributions limit and constrain the space of all possible solutions. Typical constraints for the image to be recovered, considered in the literature, are positivity, smoothness, piecewise-smoothness, or sparsity in some basis (discrete wavelet/cosine/Fourier transform, etc.). We can attempt to model these descriptions in a stochastic sense using the prior distributions. This is typically done by specifying

probabilistic relations between neighbouring pixels or their derivatives. Similar considerations may apply for the blur function.

We will consider a general exponential form of the image and blur prior distributions as follows

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\phi}) &= \frac{1}{Z_{\mathbf{x}}(\boldsymbol{\phi})} \exp\left[-U_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\phi})\right], \\ p(\mathbf{h}|\boldsymbol{\phi}) &= \frac{1}{Z_{\mathbf{h}}(\boldsymbol{\phi})} \exp\left[-U_{\mathbf{h}}(\mathbf{h}, \boldsymbol{\phi})\right], \end{aligned} \tag{3.12}$$

respectively. The normalizing terms $Z_{\mathbf{x}}(\cdot)$ and $Z_{\mathbf{h}}(\cdot)$ depend on the hyperparameters for each distribution. They may be either considered as constants if we assume known hyperparameters, or they must be calculated as $\int \exp\left[-U_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\phi})\right] d\mathbf{x}$ and $\int \exp\left[-U_{\mathbf{h}}(\mathbf{h}, \boldsymbol{\phi})\right] d\mathbf{h}$, respectively, which may cause difficulties in inference unless we assume a special form for $U(\cdot)$. In the literature, many image and blur models can be written in the form of Eq.(3.12). Below we summarize the mostly widely used prior distributions for image and blur models in the literature.

1. *Gaussian models*: The most widely used model in image deconvolution is the class of Gaussian models, given by $U_{\mathbf{x}} = \frac{\|\boldsymbol{\Delta}\mathbf{x}\|_2^2}{2\gamma_{\mathbf{x}}^2}$, then if $\det|\boldsymbol{\Delta}| \neq 0$, the term $Z_{\mathbf{x}}$ in Eq.(3.12) becomes $\left(2\pi\gamma_{\mathbf{x}}^2\right)^{N/2} \det|\boldsymbol{\Delta}|^{-1}$. The common usage of $\boldsymbol{\Delta}$ in the literature is the discrete Laplacian operator, which constrains the derivative of the image. Molina et al. [130] used this model for both image and blur, as follows

$$\begin{aligned} p(\mathbf{x}|\gamma_{\mathbf{x}}^2) &\propto \left(\gamma_{\mathbf{x}}^2\right)^{-d_{\mathbf{x}}/2} \exp\left(-\frac{\|\boldsymbol{\Delta}_1\mathbf{x}\|_2^2}{2\gamma_{\mathbf{x}}^2}\right), \\ p(\mathbf{h}|\gamma_{\mathbf{h}}^2) &\propto \left(\gamma_{\mathbf{h}}^2\right)^{-d_{\mathbf{h}}/2} \exp\left(-\frac{\|\boldsymbol{\Delta}_2\mathbf{h}\|_2^2}{2\gamma_{\mathbf{h}}^2}\right), \end{aligned} \tag{3.13}$$

   where $\boldsymbol{\Delta}_1 \in \mathbb{R}^{N\times N}$ and $\boldsymbol{\Delta}_2 \in \mathbb{R}^{M\times M}$ are two Laplacian operators, $d_{\mathbf{x}} = N - 1$ is the rank of $\boldsymbol{\Delta}_1^T\boldsymbol{\Delta}_1$, which is a singular matrix, similarly $d_{\mathbf{h}} = M - 1$.

   A similar class of Gaussian models is the Markov random field (MRF) models [199], which are widely used in image segmentation [52], image restoration [76], super resolution [162], and in image deconvolution [46, 213]. They usually promote local spatial dependences between neighbouring pixels by defining a covariance matrix that defines the relationship between neighbouring pixels. A Gibbs distribution can be defined by setting $U = \sum_{c\in C} V_c(\mathbf{x})$ in Eq.(3.12), where $V_c(\mathbf{x})$ is a potential function defined over the cliques c in the image [199]. Usually, the normalization term Z in this case is named the partition function. MRF prior distribution gives a simple way of specifying relationships between neighbouring pixels in the image. The MRF prior distribution for quadratic potentials may be written as follows

$$p(\mathbf{x}) = \frac{1}{Z}\exp\left(-\mathbf{x}^T\mathbf{B}\mathbf{x}\right) = \frac{1}{Z}\exp\left(-\sum_{c\in C}\mathbf{x}^T\mathbf{B}_c\mathbf{x}\right), \tag{3.14}$$

where $[B_c]_{i,j}$ are having a value $d_{i,j}$, which defines the relationship between the pixels i and j, which only non-zeros when the pixels are not neighbours. The normalization term now becomes $(2\pi)^{N/2} \det |\mathbf{B}|^{-1/2}$. This model is named as Gaussian Markov Random Fields (GMRF) [31, 157].

2. *Total variation prior models*: Non-quadratic image prior distributions have been also considered in the deconvolution literature by using total variation (TV) [39] regularization frameworks, with the aim of preserving edges. This prior distribution is defined in the continuous form as

$$\begin{aligned} U_{\mathbf{x}} &= \frac{1}{\gamma_{\mathbf{x}}^2} \int_{\Omega} |\bigtriangledown_1 \mathbf{x}| \, du \, dv, \\ U_{\mathbf{h}} &= \frac{1}{\gamma_{\mathbf{h}}^2} \int_{\Omega} |\bigtriangledown_2 \mathbf{h}| \, du \, dv, \end{aligned} \tag{3.15}$$

where $\bigtriangledown_1 \mathbf{x}$ and $\bigtriangledown_2 \mathbf{h}$ denote the gradients of $\mathbf{x}$ and $\mathbf{h}$ respectively, and $\Omega \in \mathbb{R}^2$ is the support of $\mathbf{x}$ and $\mathbf{h}$. The discrete forms can be written as

$$\begin{aligned} U_{\mathbf{x}} &= \frac{1}{\gamma_{\mathbf{x}}^2} \|\bigtriangledown_1 \mathbf{x}\|_1, \\ U_{\mathbf{h}} &= \frac{1}{\gamma_{\mathbf{h}}^2} \|\bigtriangledown_2 \mathbf{h}\|_1, \end{aligned} \tag{3.16}$$

where $\|\cdot\|_1$ is the ℓ_1 norm. The discrete (isotropic) definition of TV given by

$$\|\bigtriangledown \mathbf{x}\|_1 = \sum_i |\bigtriangledown_i^h \mathbf{x}| + |\bigtriangledown_i^v \mathbf{x}|, \tag{3.17}$$

where $\bigtriangledown_i^h$ and $\bigtriangledown_i^v$ are linear operators corresponding to horizontal and vertical first order differences, at pixel i, respectively. That is, $\bigtriangledown_i^h \mathbf{x} \equiv x_i - x_{i-1}$ and $\bigtriangledown_i^v \mathbf{x} \equiv x_i - x_{i-1}$.

The main difference with the GMRF model described above is that the ℓ_2 norm used in GMRF of the image gradient penalizes too much the gradients corresponding to edges and an over-smoothing effect is observed [38, 178]. In real images, object edges create sharp steps that appear as discontinuities in the intensity function. Moreover, TV prior distribution models usually begin with a formulation in the continuous image domain resulting in partial differential equations (PDEs) that must be solved. However, eventual discretization is eventually necessary [28]. Alternatively, other methods propose formulating the TV norm directly in the discrete domain [3, 21, 27, 41]. The motivation for using TV regularization for the PSF, in the literature, is due to the fact that some PSF's can have edges [39].

3. *Sparse prior models*: In image restoration, it has been shown that sparsity in some basis can serve as a good prior [24, 54]. This illustrates the fact that many types of natural images can

be sparsely represented by a dictionary of atoms. Sparse representation modelling assumes that the original image $\mathbf{x}$ can be described as $\mathbf{x} \approx \mathbf{D}\boldsymbol{\psi}$, where $\mathbf{D}$ is a dictionary and $\boldsymbol{\psi}$ is sparse. Therefore, $\mathbf{x}$ is represented by a linear combination of few columns (atoms) of $\mathbf{D}$. With these definitions, the recovery of the sparse codes given the dictionary $\mathbf{D}$ is termed sparse coding. With sparsity prior, the model in Eq.(3.3), with the representation of $\mathbf{x}$ over $\mathbf{D}$ can be written as $\mathbf{y} = \mathbf{H}\mathbf{D}\boldsymbol{\psi} + \mathbf{w}$, where an ℓ_0 prior can be assigned to $\boldsymbol{\alpha}$; e.g., $\|\boldsymbol{\psi}\|_0$. The ℓ_0-norm counts the number of non-zero entries of $\boldsymbol{\psi}$. Due to the non-convexity of the ℓ_0-norm, it is quite often relaxed to the ℓ_1-norm; $\|\boldsymbol{\psi}\|_1$ which is convex. Once $\boldsymbol{\psi}$ is estimated, the original image $\hat{\mathbf{x}}$ can be estimated as $\hat{\mathbf{x}} = \mathbf{D}\hat{\boldsymbol{\psi}}$. The dictionary $\mathbf{D}$ can be either one of the pre-defined fixed bases such as discrete cosine/wavelet/Fourier transform (DCT, DWT, DFT, respectively), etc., or can be learned from a set of training data such as K-singular value decomposition (K-SVD), method of optimal directions (MOD), and others. Sparse representation has been successfully used in various image restoration problems [58, 62, 205].

3.2.4.4 Hyperparameters prior models

The previous subsections introduced the likelihood function of the deconvolution problem as well as the choices of the prior distributions for the image to be recovered $\mathbf{x}$ and the blur function $\mathbf{h}$; $p(\mathbf{x}|\boldsymbol{\phi})$, and $p(\mathbf{h}|\boldsymbol{\phi})$ respectively. As mentioned earlier, these prior distributions might have also hyperparameters that can be fixed if they are known, or estimated if they are unknown. In this section, we study the joint estimation of the hyperparameters $\boldsymbol{\phi}$ by studying the distribution $p(\boldsymbol{\phi})$. To address this, the hierarchical Bayesian model introduces a second stage in which the $p(\boldsymbol{\phi})$ is defined, where the first stage consists of defining $p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi})$, and $p(\mathbf{x}|\boldsymbol{\phi}), p(\mathbf{h}|\boldsymbol{\phi})$.

In Bayesian literature, there has been extensive works studying the choice of the hyperparameter distribution $p(\boldsymbol{\phi})$ for which $p(\boldsymbol{\phi}, \mathbf{x}, \mathbf{h}|\mathbf{g})$ can be calculated in a straightforward way to give a tractable distribution for which Bayesian estimators can be approximated easily. These are called conjugate prior distributions [25, 149].

In addition to providing easy computation of estimators from $p(\boldsymbol{\phi}, \mathbf{x}, \mathbf{h}|\mathbf{g})$, conjugate prior distributions have also the convenient property of leading to posterior distributions (conditioned on the values of some of the other model parameters), that belong to the same family as the original prior distributions.

By taking into account conjugate prior distributions, the literature in the deconvolution and restoration considers different *a priori* models for the parameters depending on the type of the unknown parameters. Consider, the noise variance σ^2 defined in Eq.(3.7) and the regularization parameters $\gamma_{\mathbf{x}}^2$ and $\gamma_{\mathbf{h}}^2$ defined in Eq.(3.13), can be assigned inverse-Gamma distributions. For instance, for a hyperparameter ω, the PDF of the inverse-Gamma distribution defined over the support $\omega > 0$ is given by

$$p(\omega|\alpha_\omega,\beta_\omega)=\frac{{\beta_\omega}^{\alpha_\omega}}{\Gamma(\alpha_\omega)}\omega^{-\alpha_\omega-1}\exp\left(-\frac{\beta_\omega}{\omega}\right), \tag{3.18}$$

where $\alpha_\omega > 0$ and $\beta_\omega > 0$ are shape and scale parameters respectively. The inverse-Gamma distribution has the following mean and variance:

$$E(\omega)=\frac{\beta_\omega}{\alpha_\omega-1} \text{ for } \alpha_\omega>1, \quad \text{Var}(\omega)=\frac{{\beta_\omega}^2}{(\alpha_\omega-1)^2(\alpha_\omega-2)} \text{ for } \alpha_\omega>2. \tag{3.19}$$

Some authors in the deconvolution and restoration literature consider the reciprocal of a hyperparameter, and hence the assigned conjugate prior distribution would be $p(\frac{1}{\omega}) \sim \mathcal{G}(\alpha_\omega, 1/\beta_\omega)$ [21,41,130], where $\mathcal{G}$ is the Gamma distribution. Robert [153, Chap. 3, Sec. 3.3] provides a list of distributions and their corresponding conjugate ones.

The values assigned to the parameters of the conjugate prior distributions in Eq.(3.18) (i.e., α_ω and β_ω reflect the amount of knowledge known about the hyperparameter, which can yield either informative or weakly-informative prior distributions. Often most of Bayesian work in the literature assumes fixing one of the parameters of the conjugate prior distribution while introducing another level of the hierarchical Bayesian model by assigning a conjugate prior distribution to the other parameter. According to the Bayesian paradigm, uncertainty at any of these levels is incorporated into additional prior distributions [153, Chapter 10].

On the other hand, when no prior information at all is available about the hyperparameter of interest, non-informative prior distributions are considered in the Bayesian literature. Jeffreys prior distribution [85] is a widely used non-informative prior in the deconvolution and restoration literature [66, 141, 198]. It is proportional to the square root of the determinant of the Fisher information matrix, $\mathcal{I}$, which can be given as follows

$$p(\phi) \propto \sqrt{\det(\mathcal{I}(\phi))}. \tag{3.20}$$

The choice of a prior depending on Fisher information is justified by the fact that $\mathcal{I}(\phi)$ is widely accepted as an indicator of the amount of information brought by the model (or the observation) about ϕ. Hence, the values of ϕ for which $\mathcal{I}(\phi)$ is larger should be more likely for the prior distribution.

3.2.5 Bayesian Estimation for Image Deconvolution

In the previous sections of this chapter, we have studied the likelihood function as well as the prior distribution of the parameters and their corresponding hyperparameters. Bayes' rule (Equation (3.6)) is then used to compute the joint posterior distribution. The goal now is to approximate the Bayesian estimators of the unknown model parameters (e.g., $\mathbf{x}$ and $\mathbf{h}$). Depending on the prior models of the unknown parameters, finding analytical solutions to them might be difficult, hence

approximations are often considered. Point estimates, which are presented in the next subsections, of the unknown model parameters may be obtained by considering optimization algorithms, however, considering Bayesian frameworks can estimate the distributions of the parameters, which deal better with uncertainty when limited information presents [74, 91, 135]. Different inference strategies, which have been used in the deconvolution and restoration literature, are presented in the following subsections.

3.2.5.1 Maximum likelihood and maximum *a posteriori* estimators

The maximum likelihood (ML) estimation, maximizes the likelihood $p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi})$ with respect to the unknown variables

$$\{\hat{\mathbf{x}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\phi}}\}_{\text{ML}} = \underset{\mathbf{x},\mathbf{h},\boldsymbol{\phi}}{\text{maximize}} \quad p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}). \tag{3.21}$$

Note, however, that in this case, we can only estimate the hyperparameters in $\boldsymbol{\phi}$ that appear in $p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi})$ and not those who appear in $p(\mathbf{x}, \mathbf{h}|\boldsymbol{\phi})$.

Maximum *a posteriori* (MAP), on the other hand, provides point estimates by finding the values of $\mathbf{x}, \mathbf{h}$ and $\boldsymbol{\phi}$ that maximize the posterior distribution in Eq.(3.6) as follows

$$\{\hat{\mathbf{x}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\phi}}\}_{\text{MAP}} = \underset{\mathbf{x},\mathbf{h},\boldsymbol{\phi}}{\text{maximize}} \quad p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi})p(\mathbf{x}|\boldsymbol{\phi})p(\mathbf{h}|\boldsymbol{\phi})p(\boldsymbol{\phi}). \tag{3.22}$$

The ML estimation can be seen as a MAP estimation but with non-informative (flat) prior distributions on the unknown model parameters. Assuming known values for parameters is equivalent to using degenerate prior distributions. The degenerate distribution defined on $\boldsymbol{\phi}$ can be defined as follow

$$p(\boldsymbol{\phi}) = \delta(\boldsymbol{\phi}, \boldsymbol{\phi}_0) = \begin{cases} 1, & \text{if } \boldsymbol{\phi} = \boldsymbol{\phi}_0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.23}$$

The ML and MAP and estimations become

$$\{\hat{\mathbf{x}}, \hat{\mathbf{h}}\}_{\text{ML}} = \underset{\mathbf{x},\mathbf{h}}{\text{maximize}} \quad p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}_0), \tag{3.24}$$

and

$$\{\hat{\mathbf{x}}, \hat{\mathbf{h}}\}_{\text{MAP}} = \underset{\mathbf{x},\mathbf{h}}{\text{maximize}} \quad p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}_0)p(\mathbf{x}|\boldsymbol{\phi}_0)p(\mathbf{h}|\boldsymbol{\phi}_0), \tag{3.25}$$

respectively. Note that the hyperparameter vector $\boldsymbol{\phi}_0$ in this case is fixed.

Many deconvolution methods are based on (3.25). As mentioned earlier, the main differences between these methods is the form of the likelihood function, the choice of the prior distributions of the image to be recovered and/or the blur and their corresponding hyperparameters, and the optimization method used to find the solutions of the unknown parameters. Computing point

estimates of the unknown parameters using MAP estimation can be found by maximizing the joint posterior distribution in Eq.(3.5) with fixing the hyperparameter vector as in Eq.(3.25). Equivalently, this can be formulated as an optimization problem by minimizing the negative log-posterior distribution given by $\mathcal{F} = -\log[p(\mathbf{x}, \mathbf{h}|\mathbf{g})]$ with fixing the hyperparameters. This results in a minimization of the cost function in addition to a number of regularization terms, which impose soft and/or hard constraints on the solutions. The hyperparameters of the regularization terms control the trade-off between the fidelity to the data and the power of the prior distributions on the unknown model parameters. The optimization problem can be thus written as follow

$$\hat{\mathbf{x}}, \hat{\mathbf{h}} = \underset{\mathbf{x},\mathbf{h}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{g}\|_2^2 + \lambda_{\mathbf{x}} f(\mathbf{x}) + \lambda_{\mathbf{h}} f(\mathbf{h}), \tag{3.26}$$

where $f(\mathbf{x})$ and $f(\mathbf{h})$ are the image and the blur regularization functions, respectively, and $\lambda_{\mathbf{x}}$ and $\lambda_{\mathbf{h}}$ are their corresponding regularization parameters.

In the literature, the resulting optimization problem can be tackled using many algorithms. If it is differentiable, gradient-based methods, like gradient descent, (block) coordinate descent, etc. can be used [78,90,211], otherwise proximal-based methods such as the alternating direction method of multipliers (ADMM), forward-backward splitting, etc. are considered [3,8,63,107,117,133,168,210].

3.2.5.2 Sampling-based methods

The Maximum *a posteriori* estimation discussed earlier does not take into account the whole posterior probability density function. If the posterior distribution is sharply peaked at the maximum, then it does not matter as the MAP estimate is a suitable estimator; however, in the case of a heavy-tailed (broad) posterior distribution, this estimate is likely to be not a good estimator. The Minimum Mean Squared Error (MMSE) estimate attempts to find the optimal parameter values that minimize the expected mean squared error between the estimates and their true values. In practice, computing MMSE estimates analytically is difficult, though it is possible with sampling-based methods or it can be approximated using variational Bayesian methods (Section 3.2.5.4).

The most general approach to performing inference for the deconvolution problem is to simulate the posterior distribution in Eq.(3.5). This allows us to perform Bayesian inference with unknown model parameters on arbitrarily complex models in high-dimensional spaces, where no analytical solutions are available. Markov Chain Monte Carlo (MCMC) methods [153, 155] attempt to approximate the posterior distribution by generating samples that are asymptotically distributed according to the target posterior distribution using Markov chains. These samples are then used to approximate the minimum mean square error (MMSE) (or posterior mean) estimators of the unknown parameter (e.g., $\mathbf{x}, \mathbf{h}$). Formally, a Markov chain is defined by a random state $\mathbf{u}$ and a transition distribution $T(\mathbf{u}'|\mathbf{u})$ specifying the probability that a random update will go to state $\mathbf{u}'$

if it starts in state $\mathbf{u}$. Running the Markov chain means repeatedly updating the state $\mathbf{u}$ to a value $\mathbf{u}$ sampled from $T(\mathbf{u}'|\mathbf{u})$.

A conceptually simple and effective approach to building a Markov chain that asymptotically samples from the joint posterior distribution of each unknown parameter given the rest of the parameters is Gibbs sampling, in which sampling from $T(\mathbf{u}'|\mathbf{u})$ is accomplished by selecting one variable $\mathbf{u}_i$ and sampling it conditioned on its neighbours in the undirected graph. Gibbs sampler has been widely used in classical image restoration in conjunction with MRF image models [30,76]. In the deconvolution and restoration problem, if we can write down analytic expressions for the conditional distributions of all the parameters we want to estimate, given the others, we simply draw samples from each of the distributions in turn, conditioned on the most recently sampled values of the other parameters. For example if we want to simulate $p(\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}|\mathbf{g})$, the iterations would proceed as follows

Algorithm 1 Gibbs Sampling Algorithm

1: **Fixed input parameters**: Number of burn-in iterations N_{bi}, total number of iterations N_{MC}
2: **Initializations** $(k = 0)$
- Set $\mathbf{x}^{(0)}$, $\mathbf{h}^{(0)}$ and $\boldsymbol{\phi}^{(0)}$,

3: **Repeat** $(1 \leq k \leq N_{\mathbf{MC}})$
- Sample $\mathbf{x}^{(k)}$ from $p(\mathbf{x}|\mathbf{g}, \mathbf{h}^{(k-1)}, \boldsymbol{\phi}^{(k-1)})$,
- Sample $\mathbf{h}^{(k)}$ from $p(\mathbf{h}|\mathbf{g}, \mathbf{x}^{(k)}, \boldsymbol{\phi}^{(k-1)})$,
- Sample $\boldsymbol{\phi}^{(k)}$ from $p(\boldsymbol{\phi}|\mathbf{g}, \mathbf{h}^{(k)}, \mathbf{x}^{(k)})$,

4: **Set** $k = k + 1$.

The idea of Gibbs sampling is that given a multivariate distribution, it is simpler to sample from a series of conditional distributions than to marginalize by integrating over a joint distribution. Whenever it is not possible to sample from the conditional distributions, Metropolis-Hastings, can be used [214]. The idea lies on proposing a new distribution $q(\cdot|\boldsymbol{\theta})$ that is easy to sample from. The distribution we aim to sample from with density $p(\boldsymbol{\theta})$ is often called the target or objective distribution, whereas the distribution with density $q(\cdot|\boldsymbol{\theta})$ is the proposal distribution. The Markov chain corresponding to the density's parameter $\boldsymbol{\theta}$ can be constructed using the Metropolis-Hastings method by considering Algorithm 2. When the proposal distribution is symmetric, $\frac{q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(k-1)})} = 1$. Often these methods suffer from curse of dimensionality as the probability of rejection increases as a function of the dimension of the vector to be estimated.

Once a satisfied number of samples is drawn from each conditional distribution, point estimates and other statistics can be computed by Monte Carlo integration. For example, to find the MMSE estimate of the true image $\mathbf{x}$, we simply compute the mean of the samples, the posterior distribution mean or MMSE estimator of $\mathbf{x}$ can be approximated by

Algorithm 2 Metropolis-Hastings Sampling Algorithm

1: **Fixed input parameters**: Number of burn-in iterations N_{bi}, total number of iterations N_{MC}
2: **Initializations** $(k = 0)$
- Set $\boldsymbol{\theta}^{(0)}$,

3: **Repeat** $(1 \leq k \leq N_{\mathbf{MC}})$
- Draw $u \sim U(0, 1)$,
- Sample $\boldsymbol{\theta}^*$ from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(k-1)})$,
- **if** $u <$ minimum $\left(1, \frac{p(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(k-1)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(k-1)})}\right)$,
  - $\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^*$,
- **else**
  - $\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^{(k)}$,

4: **Set** $k = k + 1$.

$$\hat{\mathbf{x}} = \frac{1}{N_{\text{MC}} - N_{\text{bi}}} \sum_{t=N_{\text{bi}+1}}^{N_{\text{MC}}} \mathbf{x}^{(t)}, \tag{3.27}$$

where the samples from the first N_{bi} iterations of the sampler are discarded. These samples correspond to the transient regime or burn-in period which the Markov Chain requires to reach its equilibrium distribution, that can be determined visually from preliminary runs.

MCMC methods, unlike the optimization-based algorithms discussed in Section 3.2.5.1, do not assume the convexity of the problem. However, they are computationally expensive, and although convergence to the posterior distribution is guaranteed in theory, it can be hard to tell when this has occurred in practice. It may take a long time to explore the parameter space. Sampling methods can, for instance, be of use when it is difficult to compute the expectations of the conditional distributions by variational methods or due to the non-convexity of the problem.

3.2.5.3 Marginalizing hidden variables

The blind deconvolution inference problem can be solved by first computing estimates for $\mathbf{h}$ and $\boldsymbol{\phi}$, which can be computed as

$$\hat{\mathbf{h}}, \hat{\boldsymbol{\phi}} = \arg\max_{\mathbf{h},\boldsymbol{\phi}} \int_{\mathbf{x}} p(\mathbf{g}|\mathbf{x}, \mathbf{h}, \boldsymbol{\phi}) p(\mathbf{x}, \mathbf{h}|\boldsymbol{\phi}) p(\boldsymbol{\phi}) d\mathbf{x}, \tag{3.28}$$

then, using these estimates to compute the original image $\mathbf{x}$ as follows

$$\hat{\mathbf{x}}|_{\hat{\mathbf{h}},\hat{\boldsymbol{\phi}}} = \arg\max_{\mathbf{x}} p(\mathbf{g}|\mathbf{x}, \hat{\mathbf{h}}, \hat{\boldsymbol{\phi}}) p(\mathbf{x}|\hat{\boldsymbol{\phi}}). \tag{3.29}$$

We can also marginalize $\mathbf{h}$ and $\boldsymbol{\phi}$ to directly obtain the estimate of $\mathbf{x}$ as follow

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \int_{\mathbf{h},\boldsymbol{\phi}} p(\mathbf{g}|\mathbf{x},\mathbf{h},\boldsymbol{\phi}) p(\mathbf{x},\mathbf{h}|\boldsymbol{\phi}) p(\boldsymbol{\phi}) d\mathbf{h}\, d\boldsymbol{\phi}. \tag{3.30}$$

The two inference models provided above are called evidence-based and empirical-based analysis, respectively [128]. The marginalized variables are named hidden variables.

The expectation-maximization (EM) algorithm, first described in [51], is widely used in the deconvolution literature [94, 104] for solving ML and MAP problems, that can be regarded as having missing/hidden data. Its convergence properties to a local maxima of the likelihood or the posterior distribution is guaranteed. It is particularly suited to inverse problems in imaging. There are two main applications of the EM algorithm. The first occurs when the data indeed has missing values, due to problems with or limitations of the observation process. The second occurs when optimizing the likelihood function is analytically intractable. Each iteration of the EM algorithm consists of two processes: The E-step, and the M-step. In the expectation step, or E-step, the algorithm finds the expected value of the complete-data log-likelihood (which is easier to compute than considering the original likelihood function) with respect to the missing/hidden data and the current unknown model parameter estimates. In the maximization step, or M-step, the EM algorithm maximizes the expectation we computed in the first step. This also applies for MAP estimation by replacing the likelihood by the posterior distribution.

It is usually not possible to calculate closed forms expressions for the integrals in Eq.(3.28) and Eq.(3.30), hence approximations are considered in the deconvolution literature. For instance, Galatsanos et al. [72, 73] considered a partially known blur deconvolution problem, in which the integral in Eq.(3.28), over a distribution on $\mathbf{x}$ that is Gaussian, performed by considering a second-order Taylor expansion of $\log p(\mathbf{g}|\mathbf{x},\mathbf{h},\boldsymbol{\phi})$, then for each value of $\mathbf{h}$ and $\boldsymbol{\phi}$, the original image $\mathbf{x}$ can be computed as follow

$$\hat{\mathbf{x}}|_{\mathbf{h},\boldsymbol{\phi}} = \arg\max_{\mathbf{x}} p(\mathbf{g}|\mathbf{x},\mathbf{h},\boldsymbol{\phi}) p(\mathbf{x}|\boldsymbol{\phi}). \tag{3.31}$$

This methodology is called Laplace approximation [93, 118].

3.2.5.4 Variational Bayes

For this estimation approach, the joint posterior distribution $p(\mathbf{x},\mathbf{h},\boldsymbol{\phi}|\mathbf{g})$ can be approximated by a simpler tractable distribution $q(\mathbf{x},\mathbf{h},\boldsymbol{\phi})$ following the variational methodology [23]. The distribution $q(\mathbf{x},\mathbf{h},\boldsymbol{\phi})$ can be found by minimizing the Kullback-Leibler (KL) divergence, between the actual posterior distribution and its approximation [99] [100]. In order to obtain a tractable approximation, the family of distributions $q(\mathbf{x},\mathbf{h},\boldsymbol{\phi})$ are restricted utilizing the mean field approximation [145] so that $q(\mathbf{x},\mathbf{h},\boldsymbol{\phi}) = q(\mathbf{x})q(\mathbf{h})q(\boldsymbol{\phi})$. An iterative procedure can then be developed to estimate the distributions of the parameters $\{\mathbf{x},\mathbf{h},\boldsymbol{\phi}\}$.

The VB estimation strategy is widely used in the deconvolution and restoration literature

[18–21, 28, 129–131, 163, 192]. It is usually said that VB is what is implemented while waiting a corresponding Gibbs sampler to converge, as it provides a much faster MMSE estimation than MCMC methods.

3.2.6 Non-Bayesian Image Deconvolution

There are many other methods that are hard to cover here as the main focus in this thesis is for Bayesian models. In this section, we review some of them. Among these algorithms, there are non-parametric Deterministic Constraints Algorithms, and learning-based models.

The non-parametric Deterministic Constraints Algorithms differ from the other joint image and PSF identification methods in that they do not explicitly model the original image or the blur as stochastic or deterministic model. Rather, they consider numerical iterative formulation to impose deterministic constraints on the image to be recovered or the PSF at each step. Deterministic constraints might include positivity, energy bounds, and finite support on the image to be recovered or the PSF or both.

The iterative blind deconvolution algorithm, proposed by Ayers and Dainty [17], uses a Wiener-like iterative algorithm to estimate the image and the blur in the Fourier domain at each iteration. Moreover, it considers also positivity and finite support constraints on the image to be recovered and the PSF. The algorithm begins by initializing the image ($\hat{\mathbf{x}}^{(0)}(n)$) and the PSF ($\hat{\mathbf{h}}^{(0)}(n)$) estimates randomly, then the blur and the image are computed in the Fourier domain as in Algorithm 3.

Algorithm 3 Iterative Blind Deconvolution Algorithm

1: **Fixed input parameters**: number of iterations N_{IBA}
2: **Initializations** $k = 0$, $\mathbf{x}^{(0)}(n)$, and $\mathbf{h}^{(0)}(n)$
3: **Repeat** ($1 \leq k \leq N_{\mathbf{IBA}}$)
- Compute $\hat{\mathbf{X}}^{(k)}(p) = \text{FT}\{\hat{\mathbf{x}}^{(k)}(n)\}$,
- Compute $\tilde{\mathbf{H}}^{(k)}(p) = \frac{\mathbf{G}(p)\hat{\mathbf{X}}^{*(k)}(p)}{|\hat{\mathbf{X}}^{(k)}(p)|^2+\lambda|\hat{\mathbf{H}}^{(k-1)}(p)|^2}$,
- Compute $\tilde{\mathbf{h}}^{(k)}(n) = \text{IFT}\{\tilde{\mathbf{H}}^{(k)}(p)\}$,
- Impose positivity constraint on $\tilde{\mathbf{h}}^{(k)}(n)$ to give $\hat{\mathbf{h}}^{(k)}(n)$,
- Compute $\tilde{\mathbf{X}}^{(k)}(p) = \frac{\mathbf{G}(p)\hat{\mathbf{H}}^{*(k)}(p)}{|\hat{\mathbf{H}}^{(k)}(p)|^2+\lambda|\hat{\mathbf{X}}^{(k)}(p)|^2}$,
- Impose positivity constraint on $\tilde{\mathbf{x}}^{(k)}(n)$ to give $\hat{\mathbf{x}}^{(k+1)}(n)$,

4: **Set** $k = k + 1$.

where $\text{FT}(\cdot)$ (respectively $\text{IFT}(\cdot)$), represents the Fourier transform (respectively inverse Fourier transform) of $(\cdot)$, and $(\cdot)^*$ represents the complex conjugate of $(\cdot)$. The algorithm is run for a specified number of iterations, or until the estimates begin to converge. The final estimates depend on the constant $\lambda \in \mathbb{R}^+$, which represents the energy of the additive noise, which should be carefully chosen in order to provide reliable estimates of the original image and the blur kernel.

On the other hand, learning-based algorithms for image restoration and blind image deconvolution have been proposed in the deconvolution literature [134,144]. Such algorithms deviate from the traditional approaches in this area by utilizing prior distributions that are learned from similar images. The basic idea in such an approach is that the prior knowledge required for solving various (inverse) problems can be learned from training data, i.e., set of prototype images belonging to the same (statistical) class of images with the ones to be processed. Vector quantizer (VQ) codebooks are designed by original images and their corresponding degraded versions (by known degradation operators). During restoration, the high frequency information of a given degraded image is estimated from its low frequency information based on the learned codebooks. For blind restoration problems, a number of codebooks are designed corresponding to various versions of blurring functions. Given a blurred and noisy image, one of the codebook is chosen based on similarity measure, therefore proving identification of the PSF.

3.2.7 Why new restoration algorithms?

In this section, we introduced the Bayesian models that are widely used in the literature in image deconvolution and restoration. We noticed that no work attempted to solve the fibre core cross coupling/talk problem as an imaging inverse problem. Moreover, a parametric prior blur model for fibre core cross coupling/talk in coherent fibre bundles is missing in the literature. Also, a comparison between variational, sampling and optimization based inference strategies for restoration of irregularly spatially subsampled images is also missing. Hence, in Chapter 4, we propose a formation to the fibre core cross coupling/talk by a linear system and a Bayesian model associated with different inference strategies to estimate the unknown model parameters. This makes the restoration problem in this respect unique.

3.3 Spot Detection

3.3.1 Introduction

Quantitative analysis of microscopic biological images generally includes the detection of tiny spots (cells, bacteria, etc.). In fluorescence microscopy techniques, including endomicroscopy, the signal-to-noise ratio (SNR) can be extremely low, making automated spot detection a very challenging task, specially in the presence of background structures (elastin collagen, etc.). The objects of interest are labelled with fluorescent SmartProbes, to enhance them to fluoresce, and appear in the images as bright spots, each occupying only a few pixels. In many cases, the quality of the image data is rather low, due to limitations during the image acquisition process, which prevents the detections of these spots. This is the case in microscopic cell imaging, where excitation intensities are reduced to prevent photo-bleaching of the labelled objects, resulting in low signal-to-noise

ratio (SNR) [77]. In addition, despite recent advances in improving optical microscopic imaging techniques, the resolution of the best microscopes available is still rather lower compared to the size of sub-cellular structures, resulting in diffraction-limited appearance. As a consequence, it is sometimes difficult, even for expert clinicians and biologists, to identify objects from irrelevant background structures or noise.

In practice, automatic object-detection methods applied to fluorescence microscopy images to detect microscopic structures, reported in literature, either provide high number of false positives and/or false negatives. High false positive detections corrupt subsequent analysis with the presence of non-existent objects, and high false negative detections provides less objects than are actually present, causing subsequent analyses to be biased towards more clearly distinguishable objects. In the literature, the spot detection problem has been tackled using tracking algorithms, which consist of the detection process in the spatial domain and the linking process in the temporal domain [125]. The performance of the object tracking algorithm is crucial as poor detection would likely cause the linking procedure to yield nonsensical tracks, where correctly detected objects in one frame are connected with false detections in the next (and vice versa), or where tracks are terminated prematurely because no corresponding objects were detected in the next frame(s).

The spot detection methods reported in the literature include linear or morphological image filtering [33, 96, 170, 171], Bayesian outlier detection models [9, 53, 124] and machine learning methods [13, 87, 108, 164]. They can be classified into either unsupervised or supervised approaches. Unsupervised algorithms (linear or morphological filters and Bayesian outlier detection models) assume some object appearance model and contain parameters that need to be adjusted either manually or automatically/semi-automatically in order to get the best performance for a specific application. Supervised methods (some machine/deep learning algorithms), on the other hand, "learn" the object appearance from annotated training data usually a large number of small image patches containing only the object intensity profiles (positive samples) or irrelevant background structures (negative samples).

The rest of this chapter reviews the literature for the frequently used spot detection methods that can be applied to OEM, and is organized as follow. Section 3.3.2 provides the universal spot detection framework, and Section 3.3.3 provides the most widely used spot detection algorithms.

3.3.2 Universal Spot Detection Framework

3.3.2.1 Image formation

In fluorescence microscopy, as previously mentioned in Chapter 2, specimens are labelled with SmartProbes, which then fluoresce and is captured by a photosensitive detector (usually a CCD camera or a photomultiplier tube). The objects of interest to be detected in fluorescence microscopy application appear in images as blurred spots, which are relatively small and compact, have no clear

borders. Hence, it is better to refer to it as a "detection" problem rather than a "segmentation" one. The intensities of these objects are assumed to be higher than the background. The blurring is caused by the diffraction phenomenon and imperfections of the optical system.

Other major limitations of fluorescence microscopy are noise sources. There are two major noise sources in fluorescence microscopy, the intrinsic photon noise, which results from the random nature of photon emission, and the detector noise. Photon noise is often modelled as Poisson noise, and is independent of the detector electronics, which adds Gaussian noise. Photon noise can be reduced (which results in higher SNR), by increasing light intensity or the exposure time. However, increasing light intensity to reduce noise and improve image quality causes the photobleaching problem discussed in Chapter 2.

3.3.2.2 Spot-detection framework

Most of the spot-detection methods, considered in the literature, consist of three main steps: denoising, signal enhancement and signal thresholding [169]. Some of the steps are optional or can be combined. In step 1, in most cases, Gaussian smoothing is used for denoising, which may increase the SNR and improve image quality and object visibility for subsequent detection. In step 2 (signal enhancement), the denoised fluorescent light signal, only in the regions of the image where the actual objects exist, is enhanced, and the fluorescent signal from the background structures is suppressed. In other words, the denoised image is transformed to a new image, which is also called classification map, which does not necessarily represent the object intensity distribution any more. At this step, the value at any pixel measures the certainty in the object presence at that position. In other words, the enhanced image can also be considered as a probability map that describes the probability of object presence. In step 3 (signal thresholding), hard (binary) decision thresholds are applied in order to obtain detected objects and extract position information from the classification map. A second threshold may be required to limit the size and/or the shape of the clusters; only those clusters with size larger than $v_{\min}$ and smaller than $v_{\max}$ are considered as detected objects.

3.3.3 Detection Algorithms

In this subsection we describe the detection framework discussed above. The most characteristic part of any spot detection method is the second step of the framework (signal enhancement). As mentioned earlier in the introduction, these detection algorithms can be classified into either unsupervised (linear or morphological filtering methods and hierarchical Bayesian models) or supervised (some machine/deep learning algorithms).

3.3.3.1 Linear or Morphological filtering-based methods

1. *Laplacian of Gaussian and difference of Gaussians filters*: Laplacian filters are derivative filters used to find areas of rapid change (edges) in images. Since derivative filters are very sensitive to noise, it is common to smooth the image (e.g., using a Gaussian filter) before applying the Laplacian. This two-step process is call the Laplacian of Gaussian (LoG) operation. The 2D Laplacian of Gaussian filter can be written as

$$\mathrm{LoG}(i,j) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{i^2+j^2}{2\sigma^2}\right)\exp\left(-\frac{i^2+j^2}{2\sigma^2}\right), \tag{3.32}$$

where σ^2 is the variance of the Gaussian smoothing kernel. The LoG operator takes the second derivative of the image. Where the image is basically uniform, the LoG filter will give zero. Wherever a change occurs, the LoG will give a certain response.

In the difference of Gaussians (DoG) filter, similar to the LoG filter, the image is first smoothed by convolution with Gaussian kernel of certain width σ_1, namely $\mathbf{G}_{\sigma_1}$ resulting in

$$\mathbf{J}_{\sigma_1} = \mathbf{G}_{\sigma_1} * \mathbf{I}, \tag{3.33}$$

which is then smoothed by another smoothing Gaussian kernel of width σ_2, resulting in

$$\mathbf{J}_{\sigma_2} = \mathbf{G}_{\sigma_2} * \mathbf{I}. \tag{3.34}$$

The difference of these two Gaussian smoothed images, called difference of Gaussian (DoG), can be expressed as

$$\mathbf{S} = \mathbf{J}_{\sigma_1} - \mathbf{J}_{\sigma_2} = (\mathbf{G}_{\sigma_1} - \mathbf{G}_{\sigma_2}) * \mathbf{I}. \tag{3.35}$$

Hence the DoG filter can be defined as $\mathrm{DoG} = \mathbf{G}_{\sigma_1} - \mathbf{G}_{\sigma_2}$. As the difference between two differently low-pass filtered images, the DoG is actually a band-pass filter, which removes high frequency components representing noise, and also some low frequency components representing the homogeneous areas in the image. The frequency components in the passing band are assumed to be associated to the edges in the images.

These two filters have been used to detect blobs in the literature [81,111], however still the edges of the homogeneous regions are falsely detected as blobs.

2. *Top-hat filter*: The top-hat filtering is one of the methods used for bright spots detection in the presence of widely varying background intensities [33]. Such filters are dynamic thresholding operators. They discriminate the spots (which are groups of pixels), by their round shape and predetermined information about their intensity and size. At each pixel location in the

image to be filtered, (i, j), the average image intensity is calculated for pixels within two circular regions D_{top} and D_{brim}, denoted by $\hat{I}_{\text{top}}$ and $\hat{I}_{\text{brim}}$, respectively, defined as

$$\begin{aligned} D^{i,j}_{\text{top}} &= \{(i', j') : \sqrt{(i-i')^2 + (j-j')^2} < R_{\text{top}}\}, \\ D^{i,j}_{\text{brim}} &= \{(i', j') : R_{\text{top}} < \sqrt{(i-i')^2 + (j-j')^2} < R_{\text{brim}}\}, \end{aligned} \tag{3.36}$$

where the radius R_{top} corresponds to the "top" of the "hat" and is set to the maximum expected spot radius. The brim radius, $R_{\text{brim}} : (R_{\text{brim}} > R_{\text{top}})$, is often taken to be the shortest expected distance to the neighbouring spot. If the difference $\hat{I}_{\text{top}} - \hat{I}_{\text{brim}}$ is larger than some threshold H_{thres}, the original image intensity $\mathbf{I}(i, j)$ for that position (i, j) is copied to the classification map $\mathbf{S}$, $\mathbf{S}(i, j) = \mathbf{I}(i, j)$, otherwise $\mathbf{S}(i, j) = 0$. The procedure is repeated for each pixel, and the binary map (step 3) is obtained as $\mathbf{S}_B(i, j) = 1$ if $\mathbf{S}(i, j) \neq 0$, and $\mathbf{S}_B(i, j) = 0$ otherwise. The height H_{thres} of the top above the brim is set to the minimum intensity that a spot must rise above its immediate background. In summary, this method has only three parameters to tune manually, $(H_{\text{thres}}, R_{\text{top}}, R_{\text{brim}})$, which can be related to the object appearance. The noise reduction (step 1) in this case is implicitly done while calculating the average image intensity $\hat{I}_{\text{top}}$ and $\hat{I}_{\text{brim}}$. The averaging decreases the variance in the estimation of the noisy object and background intensity levels and improves the robustness and performance of the method.

3. *Grayscale opening top-hat filter*: Similar to the method described above, this filter uses the grayscale opening operation from mathematical morphology [175]. In order to improve the filter performance, the original image $\mathbf{I}$ is first smoothed with the Gaussian kernel with scale σ (step 1) and the grayscale opening of the denoised image with a structuring element S is done, producing the image $\mathbf{J}_S$, where a flat disk of radius r_S is commonly used in fluorescence imaging. The radius r_S is related to the size of the largest objects that we would like to detect. The top-hats are obtained after the subtraction $\mathbf{S} = \mathbf{J} - \mathbf{J}_S$ (which concludes step 2), and the whole transformation acts as a background subtraction method that leaves only compact structures smaller than the disk S, or extended objects with sufficiently narrow parts, rather than compact objects only, as does the top-hat filter. The resulting image $\mathbf{S}$ is thresholded at level ℓ_d (step 3), and then all the connected components are counted. Additional filtering with v_d can be done if the size of the connected components should be taken into account. Thus, this method has four parameters to tune, $(\sigma, r_A, \ell_d, v_d)$, all of which can be related to the object appearance.

4. *Image features-based detection*: This method is based on using some additional image information during step 2 that would help to distinguish the spots from the non uniform background, by the incorporation of local curvature information [185]. First, image denoising (step 1) is performed by a matched filter where a Gaussian kernel is used. The convolution

of the original image $\mathbf{I}$ with this kernel $\mathbf{G}_\sigma$ results in a densoised image $\mathbf{J}(i,j) = (\mathbf{G}_\sigma * \mathbf{I})(i,j)$ that exhibits high responses at the position of potential spots and low responses otherwise. The true spots in the image are characterized by uni-modal intensity distributions and a relatively high intensity. Noise and non-uniform background structures typically exhibit a random distribution of intensity changes in all directions, leading to a low local curvature. These two discriminative features (intensity and curvature) are used in combination during step 2 to create the grayscale classification map C using the denoised image (step 1) as follows

$$\mathbf{S}(i,j) = \mathbf{J}(i,j)\mathcal{K}(i,j), \tag{3.37}$$

where the curvature $\mathcal{K}(i,j)$ at each pixel of $\mathbf{J}$ is given by the determinant of the Hessian matrix $\mathbf{E}(i,j)$, where $\mathbf{E}(i,j) = (\mathbf{\Delta}.\mathbf{\Delta}^T\mathbf{I})(i,j)$ and $\mathbf{\Delta} = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$. The classification map $\mathbf{S}$ again is binarized (step 3) using the threshold ℓ_d and possibly the size threshold v_d which are not directly related to the object appearance.

3.3.3.2 Hierarchical Bayesian models

A few studies have considered hierarchical Bayesian models to solve outlier detection problems [9, 10, 53, 124]. These models offer a flexible and consistent methodology to deal with uncertainty in inference when limited amount of data or prior information is available. Moreover, other unknown parameters can be jointly estimated within the algorithm such as noise variance(s) and regularization parameters. As such, they represent an attractive way to tackle ill-posed inverse problems. These methods rely on selecting an appropriate prior distribution for the unknown image and remaining unknown parameters. The full posterior distribution can then be derived using the Bayes' rule, and then exploited by optimization or simulation-based (Markov chain Monte Carlo) methods.

Altmann et al. [9, 10] proposed a hierarchical Bayesian algorithm for linear spectral unmixing of hyperspectral images that accounts for anomalies present in the data. The model proposed assumes that the pixel reflectances are linear mixtures of unknown endmembers (spectral signatures), corrupted by an additional non-linear term modelling anomalies and additive Gaussian noise. A Markov random field is used for anomaly detection based on the spatial and spectral structures of the anomalies. This allows outliers to be identified in particular regions and wavelengths of the data cube. A Bayesian algorithm was proposed to estimate the parameters involved in the model, using MCMC sampling methods, in which a Gibbs sampler was used, yielding a joint linear unmixing and anomaly detection algorithm. The algorithm was applied for spectral unmixing and anomaly detection in multispectral remote sensing datasets.

McCool et al. [124] addressed the problem of joint image denoising and outlier detection in irregularly spatially subsampled images. The problem is formulated by assuming that the observa-

tions are a linear combination of original irregularly spaced samples which is corrupted by additive observation noise and sparse outliers. This problem is addressed within a hierarchical Bayesian framework, where suitable prior distributions are assigned to the unknown model parameters. A Gaussian Markov random field prior distribution is considered for the original intensities to be estimated as it can capture the spatial correlation between neighbouring samples. A binary labels are used to locate the spatial position of the sparse outliers. A Markov Chain Monte Carlo method which is implemented using a Gibbs sampler is then used to perform Bayesian inference using the posterior distribution associated with the resulting Bayesian model. Simulations conducted on simulated data showed promising results of the proposed method in terms of data restoration and outlier detection.

3.3.3.3 Machine/Deep learning-based methods

In machine/deep learning approaches [11, 13, 48, 87, 108, 164, 203, 204], the dataset is usually split into training, testing and validation sets. In the training phase, a set of features are extracted that can capture discriminative structures of the objects to be detected. A model is then trained by pairing inputs with their expected outputs which are also known as ground truth. This trained model is then used to estimate the output of the test dataset. The validation dataset is often used to tune the model hyperparameters.

As an example in machine learning, Seth et al. [164] considered a supervised learning approach to detect bacteria and cells on fibred confocal fluorescence microscopy images. A database of annotated frames were created, where a clinician has dot-annotated bacteria or cells. The images are first denoised (step 1) by setting the lowest intensity values below 1% quantile to 0, and the intensity values above 99% quantile to the respective 99% quantile values. Then features (intensity values) are extracted by considering patches around each pixel, and train a classifier to predict if a bacterium or cell is present at that pixel. They suggested using a multi-resolution spatio-temporal template matching scheme using radial basis functions. Spatio-temporal analysis allows better capturing of the "blinking" effect of a bacterium, whereas multi-resolution analysis allows better discrimination between bacterial dots or cellular structure and elastin background. The classifier returns a probability value at each pixel. These probability values are then thresholded (step 3), and pixels that exceed this threshold value are counted after non-maximum suppression to estimate bacterial load.

On the other hand, several studies have considered neural networks to solve the outlier detection problem [48, 203, 204]. For instance, in [204], the authors proposed a convolutional neural network (CNN) with a compressive sensing (CS) scheme in order to detect cells in optical microscopy images. This method, referred to as CNN-CS, uses encoding of the output pixel space which is the sparsely labelled pixel locations indicating cell centers. They employ random projections to encode the output space to a compressed vector of fixed dimension. Then, CNN regresses this

compressed vector from the input pixels. Furthermore, sparse cell locations on the output pixel space are recovered from the predicted compressed vector using ℓ_1-norm optimization.

In [48], the authors proposed a method to detect mitosis by a supervised deep neural network (DNN) as a pixel classifier. The DNN is a max-pooling CNN, which directly operates on raw RGB data sampled from a square patch of the source image, centred on the pixel itself. The DNN is trained to differentiate patches with a mitotic nucleus close to the center from all other windows. Mitosis in unseen images are detected by applying the classifier on a sliding window.

In [203], the authors proposed a cell counting framework based on fully convolutional regression networks. The ground truth is provided as dot-annotations, where each dot corresponds to one cell. For training, the dot annotations are each represented by a Gaussian, and a density surface $\mathbf{D}(n)$ is formed by the superposition of these Gaussians. The task is to regress this density surface from the corresponding cell image $\mathbf{I}(n)$. This is achieved by training a CNN using the mean square error between the output heat map and the target density surface as the loss function for regression. At inference time, given an input cell image $\mathbf{I}(n)$, the CNN then predicts the density heat map $\mathbf{D}(n)$.

3.3.4 Why new spot detection algorithms?

It can be noted that no unsupervised algorithms are considered in the literature for detection of fluorescently labelled bacteria in optical endomicroscopy images. Moreover, the only introduced supervised-based algorithm for bacterial detection in OEM datasets [164] is not fully automatic in the sense that it requires the user to tune a set of parameters to obtain the final bacteria locations, which might not be robust should bacteria concentration and brightness change amongst different frames/videos. Hence, there is a clear unmet need for a fully automatic algorithm that is adaptive for different bacteria concentrations in the different frames of the same dataset. Moreover, we noticed the lack of dictionary learning-based algorithms which have been found to be efficient in various imaging inverse problems [58, 156, 186].

3.4 Summary

In this chapter, we reviewed the most widely used methods for image deconvolution, restoration and spot detection. For image deconvolution and restoration, the problem is often formulated as a linear inverse problem in which the observations are assumed to result from a convolution acting on the original image and corrupted by additive observation noise. In order to obtain estimations for the unknown model parameters, we focused on Bayesian models, where all model parameters and observed variables are assumed to be unknown stochastic quantities, by assigning them probability distributions based on prior beliefs. The Bayesian paradigm first defines the likelihood function based on noise statistics, then prior distributions are assigned to unknown model parameters and hyperparameters, the full posterior distribution can then be computed using

Bayes' rule, and Bayesian inference can be performed in order to provide estimates for the unknown model parameters and may be hyperaparametrs. Deconvolution and restoration problems were classified into either *a priori* known blurs, at which the original image can be estimated assuming the blur is known, where in this case can be estimated by experimental approaches, or blind deconvolution and restoration problems at which both the true image and the blur are unknowns. We then provided the widely used image and blur prior models provided in the deconvolution and restoration literature. Then a set of Bayesian inference methods that are widely used in the deconvolution and restoration literature were introduced. These methods either compute the maximum *a posteriori* (MAP) estimate, maximum likelihood (ML) estimate or minimum mean square error (MMSE) estimate. The mostly widely used optimization method for providing MAP estimate is the ADMM method, in which, the hyperparameter vector is fixed. The expectation maximization algorithm was also used for both MAP and ML estimations. MMSE, on the other hand, can be obtained by either sampling-based methods, which generates samples distributed according to the target posterior distribution using Markov chain Monte Carlo (MCMC) methods, or approximating the posterior distribution using variational Bayes method. These two approaches also allow the estimation of the hyperparameters associated with the prior distributions. Then a set of deconvolution and restoration methods that can not be obtained from Bayesian paradigm were briefly reviewed.

In the second part of this chapter, we reviewed the literature on the most widely used methods for spot detection. Most of the revised methods include three main steps; denoising, enhancement and thresholding. These methods can be classified into either unsupervised or supervised based on the second step (enhancement). Unsupervised algorithms, like morphological filters and Bayesian outlier detection models, assume some object appearance model and adjust parameters in order to get the best performance. Supervised methods like machine learning algorithms, on the other hand, learn the object appearance from annotated training data, and then use this learned model in order to provide spot locations on testing data.

The next chapter (Chapter 4) explains how we formulated the cross coupling and field of view pixelation problems into a deconvolution and restoration task and the model proposed to solve these problem. Chapters 5 and 6 provide our proposed methods for endomicroscopic detection of fluorescently labelled bacteria.

Chapter 4

Deconvolution and Restoration of Optical Endomicroscopy Images

This chapter has been adapted from the journal paper [Eldaly et al. (2018a)] and the conference paper [Eldaly et al. (2018c)] (both published).

Contents

4.1 Introduction

As we previously mentioned in Chapter 2, optical endomicroscopy (OEM) employs a proximal light source, laser scanning or Light Emitting Diode (LED) illumination, linked to a flexible fibre bundle, performing microscopic fluorescent imaging at its distal end. Probe-based confocal laser endomicroscopy, is currently the most widely used clinical OEM platform approved for clinical use. However, novel and low-cost OEM architectures employing wide-field LED illumination sources has been introduced [82, 147, 165]. Wide-field fibre optic imaging devices, such as the one being developed by our group (Proteus, Edinburgh-UK) provide sparse and usually irregularly-spaced intensity readings of the scene, due to the irregular packing of the fibre cores within the fibre bundle. Fibre bundles usually contain approximately $25,000$ fibre cores that are transmitting and collecting the light simultaneously. Note that it is only the fibre cores which contain information while the cladding, (the space between the fibre cores), does not.

One of the main challenges of OEM images is the restoration of the signals at the receiver for better image visualization and/or subsequent analysis. Fiber core cross coupling and the pixellation effect produced by the irregular packing of fibre cores are the main reasons for image degradation in this type of imaging [152, 201]. Perperidis et al. [146] have quantified the average spread of inter-core coupled light, with approximately a third of the overall light coupling to neighbouring cores (see chapter 3). Consequently, cross coupling causes severe blurring in the resulting images, whose restoration is formulated as an inverse problem.

In this chapter, we address the problem of cross coupling between fibre cores and the sparse sampling by imaging fibre bundles (pixellation effect), which are formulated as a deconvolution and restoration problem. We propose a hierarchical Bayesian model to solve this problem and compare three estimation algorithms to exploit the resulting joint posterior distribution. The first

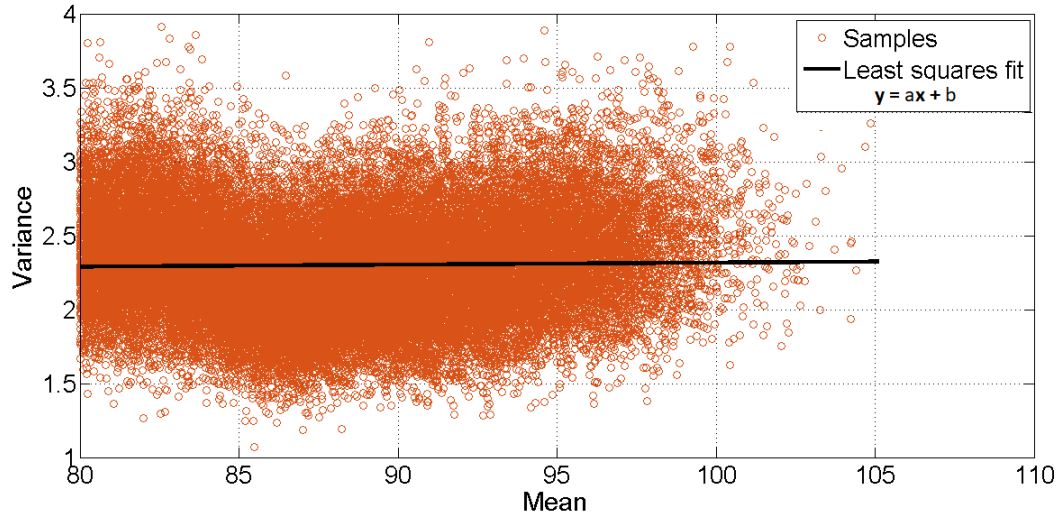
method is based on Markov chain Monte Carlo (MCMC) methods, however, it exhibits a relatively long computational time. The second and third algorithms deal with this issue and are based on a variational Bayes (VB) approach and MAP estimation via an alternating direction method of multipliers (ADMM) algorithm respectively.

Noise model

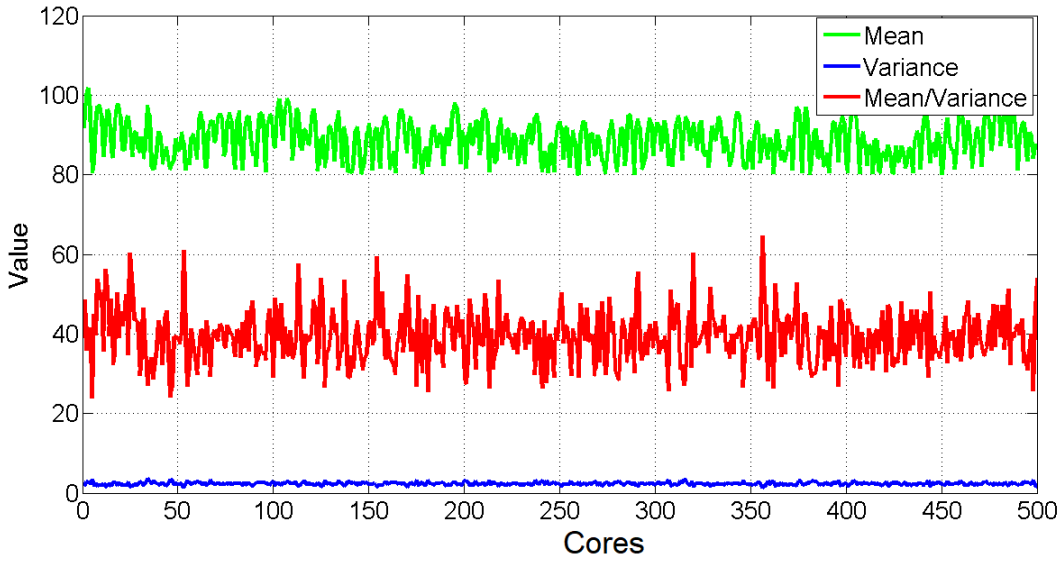
In wide-field endomicroscopy, the entire specimen is flooded with light from light sources. All parts of the specimen in the optical path are excited at the same time and the resulting fluorescence is detected by the endomicroscope's camera. In contrast, a confocal endomicroscope uses point illumination and a pinhole in an optically conjugate plane in front of the detector to eliminate out-of-focus signal. As much of the light from sample fluorescence is blocked at the pinhole, this increased resolution is at the cost of decreased signal intensity, and hence the Poissonian noise model is very common in confocal endomicroscopy [160]. In contrast, wide-field endomicroscopy acquires more fibre fluorescence and more tissue background fluorescence than confocal endomicroscopy thus increasing the effective number of photon counts > 500 per pixel when imaging tissue, which can be modelled by an independent and identically distributed (i.i.d.) Gaussian noise model [36,148,160]. This assumption was considered after a detailed analysis of a homogeneous region of interest (ROI) of a sequence of 90 background images obtained by the wide-field optical endomicroscopy system. In the images we analysed, it seems that the observation noise includes, in addition to shot noise arising from detection of photons, additional sources of randomness, e.g., caused by signal amplification, offset and discretization. Moreover, it seems that at the centre of each core, the impact of such sources of uncertainty is more important than the impact of shot noise. Indeed, while shot noise generally induces correlation between the signal mean and its variance (Poisson noise), in practice we observed negligible changes in intensity variance across the different cores. Note however, that the results might be different for other settings (e.g, much lower photon counts and higher amplification). For such scenarios, more complex noise models should be considered, including mixtures of Gaussian and scaled Poisson models [47] with deconvolution methods. They would however require a careful characterization of the image from the detected photons. The noise at central fibre core intensities was characterized by subtracting the mean signal from the individual signal realizations, and measuring the variance of the residual noise as a function of the mean signal strength. Fig. 4.1 illustrates the Gaussian distribution of the noise as the variance of the noise is core independent and constant on the mean signal strength, which is well captured by a zero mean Gaussian noise distribution.

Consequently, in this work, we consider a noisy observation vector $\mathbf{g}$, of an original intensity vector $\mathbf{x}$, that is modelled by the following linear forward model

$$\mathbf{g} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{4.1}$$



(a)



(b)

Figure 4.1 (a) Variance versus mean signal strength for fibre core intestines in one of the ROIs, which is of size 300×500, (b) plot of mean, variance, and the ratio between mean and variance for fibre core intensities in the processed ROI.

where $\mathbf{A}$ is the matrix representing a linear operator which can model different degradation. Here, $\mathbf{A}$ models fibre core cross coupling and/or spatial blur. We specify the dimensions of the variables later in the text. In Eq.(4.1), the vector $\mathbf{w}$ stands for additive noise, modelling observation noise and model mismatch and is assumed to be a white Gaussian noise sequence, as we demonstrated previously.

As previously mentioned in Chapter 3, the problem of estimating $\mathbf{x}$ from $\mathbf{g}$ is an ill-posed linear inverse problem (LIP); i.e., the matrix $\mathbf{A}$ is singular or very ill-conditioned. Consequently, this problem requires additional regularization (or prior information, in Bayesian inference terms) in order to reduce uncertainties and improve estimation performance.

Contributions

The main contributions of this work are fivefold:

1. We address the problem of deconvolution and restoration in OEM. To the best of our knowledge, it is the first time this problem is addressed in a statistical framework by using a hierarchical Bayesian model.

2. We develop algorithms dedicated to irregularly sampled images which do not rely on strong assumptions about the spatial structure of the sampling patterns. The developed methods can thus be applied to a wide range of imaging systems, and fibre bundle designs.

3. We provide new parametric blur model for the fibre core cross coupling/talk problem in coherent fibre bundles.

4. We derive three estimation algorithms associated with the proposed hierarchical Bayesian model and compare them using extensive simulations conducted using controlled and real data. The first algorithm generates samples distributed according to the posterior distribution using Markov chain Monte Carlo (MCMC) methods [154]. This approach also allows the estimation of the hyperparameters associated with the prior distributions. However, as mentioned previously, the resulting MCMC-based algorithm presents a high computational complexity. The second and third algorithms deal with this limitation and approximate the joint posterior distribution. The second algorithm uses the variational Bayes (VB) methodology [23, 119] to approximate the joint posterior distribution by minimizing the Kullback–Leibler (KL) divergence between the true posterior distribution and its approximation [99]. It can also estimate the hyperparameters associated with the prior distributions, and hence it is totally automatic, as is the MCMC-based method. The third algorithm is based on the alternating direction method of multipliers (ADMM). Although the low computation complexity of this algorithm, the hyperparameters associated with the

prior distributions need to be chosen carefully by the user, and hence it is considered as a semi-supervised method.

5. We use Gaussian Processes (GP) to interpolate the resulting samples to provide a meaningful image and quantify uncertainties at each interpolated sample, which is the first time to be applied to OEM.

The rest of this chapter is organized as follows. Section 4.2 discusses the cross coupling problem and formulates the problem of deconvolution and restoration of OEM data. The proposed hierarchical Bayesian model is then presented in Section 4.3. Section 4.4 introduces the three proposed estimation algorithms based on MCMC and optimization. Results of simulations conducted using synthetic and real datasets are discussed in Section 4.6 and Section 4.7, respectively. Conclusions and future work are finally reported in Section 4.8.

4.2 Problem Statement: Deconvolution of OEM Images

Fig. 4.2 illustrates what happens in the fibre bundle when receiving fluorescent light from an object being imaged. The vectors $\mathbf{x}$, and $\mathbf{g}$ represent light intensities at the object being imaged, which is also the distal end of the fibre bundle, and at the image plane respectively. The transform $\mathbf{H}$ represents the cross coupling effect defined later in the text, $\mathbf{C}$ represents the spatial blur acting between the proximal end of the fibre bundle and the image plane, whereas $\mathbf{C}'$ is that between the distal end of the fibre bundle and the tissue being imaged. The two spatial blurs $\mathbf{C}$ and $\mathbf{C}'$ are spatially variant, $\mathbf{C}$ can be characterized as the distance d between the image plane and the proximal end of the fibre is known, whereas $\mathbf{C}'$ cannot be fully characterized as the distance between the distal end of the fibre and the tissue being imaged is unknown, and the frames here are analyzed independently. The aim here is to recover the intensity vector $\mathbf{x}$, assuming that the spatial blur $\mathbf{C}'$ is negligible.

Fig. 4.3 provides an illustrative example of cross coupling between fibre cores. If an individual fibre core is illuminated in $\mathbf{x}$, the neighbouring cores in $\mathbf{g}$ will be affected by a specific percentage of the incident light on the illuminated core. Experimental results in current fibre bundle (which might be different for other bundles) showed that around 61% of the light transmitted through a single core remains in that core, around 34% migrates to the immediate neighbouring cores, around 4% to the second order neighbours and less than 1% to the third, fourth, and fifth order neighbours [146].

Fig. 4.4 illustrates how we construct the forward observation model to mimic the same output as the endomicroscopy imaging system. The first image on the left-hand side of the figure represents the illumination of one fibre core. This results in cross coupling to the neighbouring cores (convolution with a first linear operator $\mathbf{H}$), then the spatial blurring effect around each fibre core

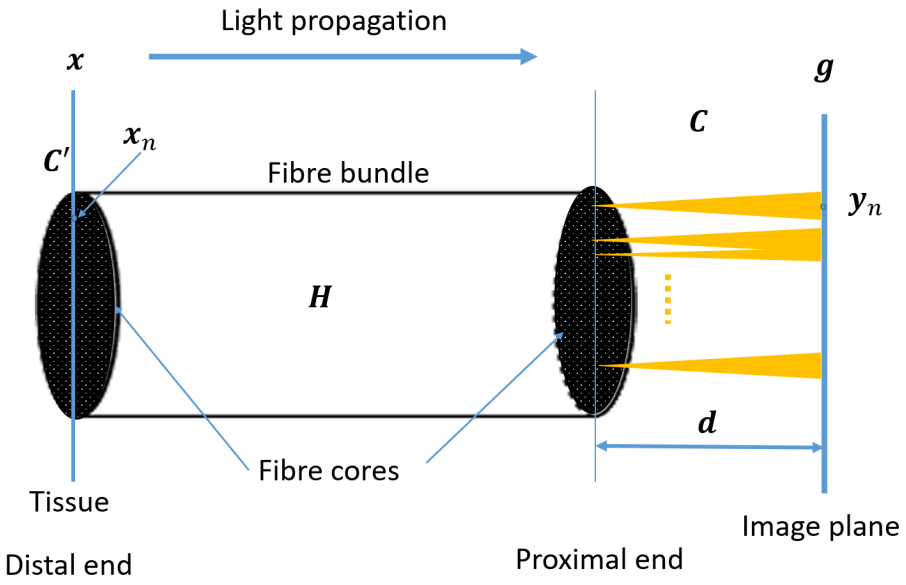


Figure 4.2 Schematic diagram showing the forward model in 2D-OEM.

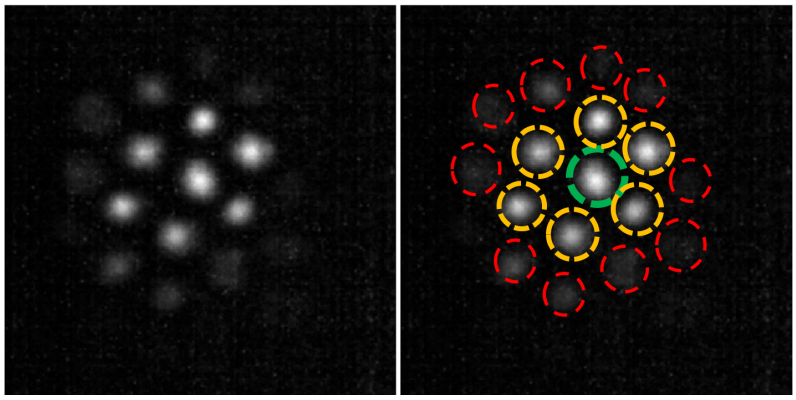


Figure 4.3 Example of cross coupling between fibre cores, the green circle represents the central illuminated core and the yellow and red ones represent the immediate and further neighbours respectively.

(convolution with a second linear operator **C**) and finally the fourth image of the figure shows the final system output after adding white Gaussian noise.

The linear model in Eq.(4.1) can now be written as

$$\mathbf{g} = \mathbf{CHx} + \mathbf{w}, \tag{4.2}$$

where **A** in Eq.(4.1) is replaced by **CH** in Eq.(4.2), the vector **g** is the observed data matrix, and **x** is the image to be restored.

Fibre core cross coupling has been identified and quantified experimentally in coherent fibre bundles [**Perperidis et al. 2017**], hence in this work, we choose to use *a priori*-identified blur methods to model the blur function which would reduce the computation cost of the estimation

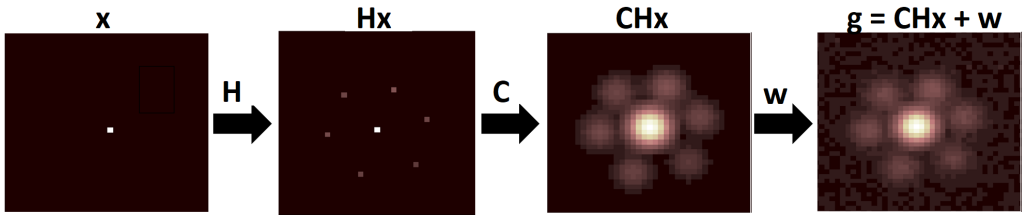


Figure 4.4 Representation of the endomicroscopy system forward model.

algorithms due to the reduction in number of variables need to be estimated. From preliminary results, we propose to model cross-coupling by an isotropic zero mean 2D generalized Gaussian kernel applied to the fibre intensities as follows

$$[\mathbf{H}]_{i,j} = \exp\left(-\left(\frac{d_{i,j}}{\alpha_{\mathbf{H}}}\right)^{\beta_{\mathbf{H}}}\right), \tag{4.3}$$

where $d_{i,j}$ denotes the euclidean distance between the cores (or spatial locations) i and j, which corresponds to approximately 3.3 pixels between neighbouring cores. From Eq.(4.3), it can be seen that neighbouring fibre cores will be more closely coupled than distant ones. The values of $\alpha_{\boldsymbol{H}}$ and $\beta_{\boldsymbol{H}}$, which control the amount of cross-coupling (the higher, the more coupling) and which are system dependent, are adjusted from preliminary measurements (calibration). Note that other cross-coupling models could also be considered instead of Eq.(4.3) depending on the imaging system used. Moreover, as previously mentioned in Chapter 3, background offset and gain [161] arising due to the difference in fibre core sizes are neglected here, in order to make the estimation of the unknown model parameters tractable.

The spatial blur affecting each fibre core can be modelled by a Gaussian spatial filter, as illustrated in Fig. 4.5, which shows a background image i.e., an image from a sample presenting constant intensity, using an endomicroscopy imaging system, and a zoomed-in region of this image, bright and dark areas represent fibre cores and their cladding, respectively. The intensity profile across one line in this image is a series of Gaussian kernels. However, the variation of the shape and width of the kernels is due to the variation in core sizes.

Due to the variation in core sizes, the blurring kernel $\mathbf{C}$ varies, and hence the core responses tend to overlap. So the complete model in Eq.(4.2) becomes more complex, and potentially computationally expensive for long image sequences (videos). Indeed, there is no structure in $\mathbf{C}$ which allows us to compute $\mathbf{CHx}$ rapidly. Hence we propose a simplification of this model and represent each core by a single intensity value. The mean intensities of fibre core pixels could be used, but the overlap between the cores makes its computation difficult. Since the variation of the width of this blur is not too significant, the maximum intensity of each core is considered instead ($\mathbf{y}_n$ in Fig. 4.2).

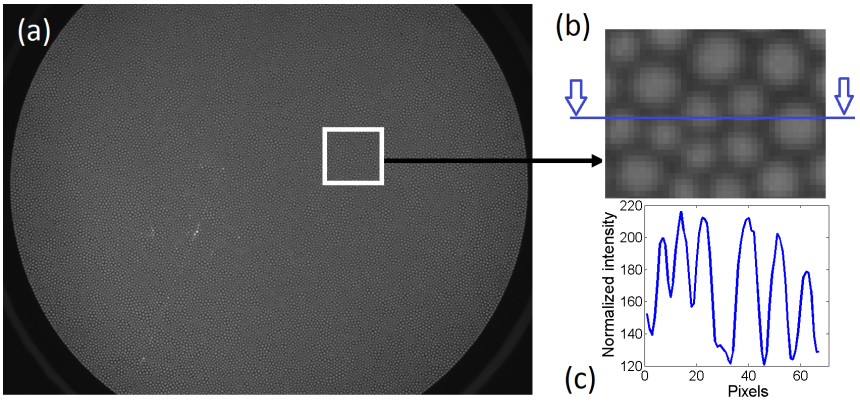


Figure 4.5 (a) A background image, (b) a zoomed part of the image, and (c) the intensity profile across one line in the image.

Following the above mentioned points, the model in Eq.(4.2) can be reduced to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}. \tag{4.4}$$

Assume that N is the total number of pixels in the image, and N_1 represents the number of fibre cores in the image, the input $\mathbf{y} \approx \mathbf{C}^{+}\mathbf{g} \in \mathbb{R}^{N_1}$, where $\mathbf{C}^{+}$ is the pseudo-inverse of $\mathbf{C}$, and the output $\mathbf{x} \in \mathbb{R}^{N_1}$ are two vectors representing central core intensities, where, $N_1 \ll N$, and $\mathbf{H} \in \mathbb{R}^{N_1 \times N_1}$. The noise $\mathbf{w} \in \mathbb{R}^{N_1}$ is assumed to be additive white noise which is i.i.d. zero mean Gaussian noise with variance σ^2, denoted as $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, where $\sim$ means "is distributed according to" and $\mathbf{I}$ is the identity matrix. Here, we consider recovering only fibre core intensities rather than a full image because of the overlap of fibre cores due to their different sizes, which makes the problem difficult to formulate as mentioned earlier. Moreover, the number of fibre cores forming the image is much smaller than the number of pixels ($N_1 \ll N$), hence we will benefit from the reduction of the size of processed data.

The problem investigated in this work is to estimate the actual intensity values $\mathbf{x}$, and the noise variance σ^2 from the observation vector $\mathbf{y}$. As mentioned previously, to solve this problem, we propose a hierarchical Bayesian model and a set of different estimation methods to estimate the unknown parameters.

4.3 Hierarchical Bayesian Model

This section introduces a hierarchical Bayesian model to estimate the unknown parameter vector $\mathbf{x}$ and σ^2. This model is based on the likelihood function of the observations and on prior distributions assigned to the unknown parameters.

4.3.1 Likelihood

Eq. (4.4) yields that $\mathbf{y}|(\mathbf{x}, \sigma^2) \sim \mathcal{N}(\mathbf{Hx}, \sigma^2\mathbf{I})$. Consequently, the likelihood can be expressed as

$$f(\mathbf{y}|\mathbf{x}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{N_1/2} \exp\left(-\frac{\|\mathbf{y} - \mathbf{Hx}\|_2^2}{2\sigma^2}\right). \tag{4.5}$$

4.3.2 Parameter Priors

4.3.2.1 Prior for the underlying intensity field x

A truncated multivariate Gaussian distribution (MVG) is assigned to the intensity field $\mathbf{x}$.

$$f(\mathbf{x}|\gamma^2) \propto \left(\gamma^2\right)^{-d/2} \exp\left(-\frac{\mathbf{x}^T\boldsymbol{\Delta}^{-1}\mathbf{x}}{2\gamma^2}\right) 1_{\mathbb{R}^+}(\mathbf{x}), \tag{4.6}$$

where $1_{\mathbb{R}^+}(\mathbf{x})$ is the indicator function defined on the positive set of $\mathbf{x}$, γ^2 controls the global correlation between intensities, and the covariance matrix $\boldsymbol{\Delta}$ which defines the spatial correlation between the cores is defined by

$$[\boldsymbol{\Delta}]_{n,n'} = \exp\left(-\left(\frac{d_{n,n'}}{\ell}\right)^{\kappa}\right), \tag{4.7}$$

where $d_{n,n'}$ denotes the distance between the spatial locations n and n', and $d = N_1$. Equations (4.6) and (4.7) promote smooth intensity variations between neighbours while ensuring that the prior dependence between neighbouring cores decrease as $d_{n,n'}$ increases. In this work $d_{n,n'}$ is the standard euclidean distance. The parameters ℓ, κ were learned from the irregular sampling pattern of the OEM system. Precisely, we used known images and selected (ℓ, κ) by maximum likelihood estimation, which occurs when $p(\ell, \kappa|\mathbf{x})$ is at its greatest, which corresponds to maximizing $\log p(\ell, \kappa|\mathbf{x})$. While γ^2 is left unknown for each image, (ℓ, κ) are fixed in the rest of the simulations as the average values obtained with the training images.

Considering such a prior distribution is equivalent to assuming a Gaussian process on $\mathbf{x}$. This allows us to interpolate the resulting deconvolved intensities using Gaussian processes [150] as we will see in section 4.5.

4.3.2.2 Prior for the noise variance σ^2

As earlier mentioned in chapter 3, the hyperparameters can be assigned conjugate distributions, for which their conditional distribution can be calculated in a straightforward way to give a tractable distribution where Bayesian estimators can be approximated. Hence, in this work, a conjugate inverse-Gamma $\mathcal{IG}$ prior distribution is assigned to the noise variance σ^2

$$f(\sigma^2|\alpha, \beta) \sim \mathcal{IG}(\alpha, \beta), \tag{4.8}$$

where $\alpha = 10$ is fixed arbitrarily, while the hyperparameter β is estimated within the algorithm.

4.3.2.3 Prior for the hyperparameter β

The hyperparameter associated with the parameter prior distribution defined above is assigned to a conjugate Gamma distribution:

$$\beta \sim \mathcal{G}(\alpha_o, \beta_o), \tag{4.9}$$

where α_o and β_o are fixed and user-defined parameters which might depend on the quality of the data to be recovered. In this work, we fixed $(\alpha_o, \beta_o) = (10, 0.1)$ arbitrarily.

4.3.2.4 Prior for the hyperparameter γ^2

To reflect the lack of prior knowledge about the regularization parameter γ^2 in Eq.(4.6), the following weakly informative conjugate inverse-Gamma prior distribution is assigned to it.

$$\gamma^2 \sim \mathcal{IG}(\eta, \nu), \tag{4.10}$$

where (η, ν) are fixed to $(\eta, \nu) = (10^{-3}, 10^{-3})$. Note that we did not observe significant change in the results when changing these hyperparameters.

The next section derives the joint posterior distribution of the unknown parameters associated with the proposed Bayesian model.

4.3.3 Joint posterior distribution

Assuming the parameters $\mathbf{x}$ and σ^2 are *a priori* independent, the joint posterior distribution of the parameter vector $\boldsymbol{\Omega} = \{\mathbf{x}, \sigma^2\}$ and hyperparameters $\boldsymbol{\phi} = \{\beta, \gamma^2\}$ can be expressed as

$$f(\boldsymbol{\Omega}, \boldsymbol{\phi}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\Omega}) f(\boldsymbol{\Omega}|\boldsymbol{\phi}) f(\boldsymbol{\phi}), \tag{4.11}$$

where

$$f(\boldsymbol{\Omega}|\boldsymbol{\phi}) = f(\mathbf{x}|\gamma^2) f(\sigma^2|\beta), \text{ and } f(\boldsymbol{\phi}) = f(\gamma^2) f(\beta). \tag{4.12}$$

The directed acyclic graph (DAG) summarizing the structure of proposed Bayesian model is depicted in Fig. 4.6. This posterior distribution will be used to evaluate Bayesian estimators of $\boldsymbol{\Theta} = \{\boldsymbol{\Omega}, \boldsymbol{\phi}\}$. For this purpose, we propose three algorithms: an MCMC-based approach and two optimization-based approaches, in which VB and ADMM are considered. The first approach uses an MCMC method to evaluate the minimum-mean-square-error (MMSE) estimator of $\boldsymbol{\Theta}$ by generating samples according to the joint posterior distribution. Moreover, it allows the estimation of the hyperparameter vector $\boldsymbol{\phi}$ along with the noise variance σ^2. However, it exhibits a relatively long computational time. The second and third algorithms which deal with this issue and provide faster MMSE estimate for the VB approach and MAP estimate for the ADMM approach. The VB

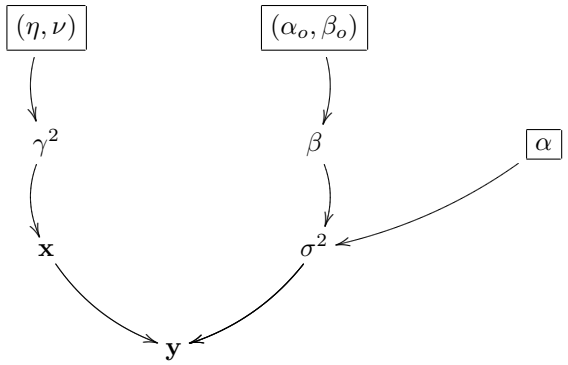


Figure 4.6 Graphical model for the proposed hierarchical Bayesian model (fixed quantities appear in boxes).

approach approximates the joint posterior distribution in Eq.(4.11) by minimizing the Kullback-Leibler (KL) divergence between the true posterior distribution and its approximation [99]. The ADMM approach is achieved by maximizing the posterior distribution Eq.(4.11) with respect to (w.r.t.) $\boldsymbol{\Theta}$. Note however, that the hyperparameters $\boldsymbol{\phi}$ as well as σ^2 are fixed for this approach. The three estimation algorithms are described in the next section.

4.4 Bayesian Inference

4.4.1 MCMC algorithm

To overcome the challenging derivation of Bayesian estimators associated with $f(\boldsymbol{\Theta}|\mathbf{y})$, we propose to use an efficient MCMC method to generate samples asymptotically distributed according to the posterior distribution presented in Eq.(4.11). More precisely, we consider a Gibbs sampler described next. The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest [[154], Chap. 10]. In this work, we propose to sample sequentially the elements of $\boldsymbol{\Theta}$ using updates that are detailed below (see [75] for how to compute the conditional distributions of model parameters).

4.4.1.1 Sampling the intensity field x

From Eq.(4.11), since the prior distribution Eq.(4.6) is conjugate to the Gaussian distribution, the full conditional distribution of $\mathbf{x}$ is given by

$$f(\mathbf{x}|\mathbf{y}, \sigma^2) \sim \mathcal{N}_{\mathbb{R}^+}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{4.13}$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \sigma^{-2}\boldsymbol{\Sigma}^T\mathbf{H}^T\mathbf{y}, \\ \boldsymbol{\Sigma} &= \left(\sigma^{-2}\mathbf{H}^T\mathbf{H} + \gamma^{-2}\boldsymbol{\Delta}^{-1}\right)^{-1}. \end{aligned} \tag{4.14}$$

Sampling from Eq.(4.13) can be achieved efficiently by using the Hamiltonian method proposed in [142].

4.4.1.2 Sampling the noise variance σ^2

By cancelling out the terms that do not depend on σ^2 in the posterior distribution in Eq.(4.11), its conditional distribution can be written as

$$f(\sigma^2|\mathbf{y},\mathbf{x}) \sim \mathcal{IG}\left(\alpha + \frac{N_1}{2}, \beta + \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2}{2}\right), \tag{4.15}$$

which is easy to sample from.

4.4.1.3 Sampling the hyperparameters β and γ^2

It can be easily shown that β can be sampled from the following Gamma distribution

$$f(\beta|\sigma^2) \sim \mathcal{G}\left(\alpha + \alpha_o, \frac{\sigma^2\beta_o}{\sigma^2+\beta_o}\right). \tag{4.16}$$

In a similar fashion to the noise variance, γ^2 can be sampled from the following inverse-Gamma distribution

$$f(\gamma^2|\mathbf{x}) \sim \mathcal{IG}\left(\eta + \frac{N_1}{2}, \nu + \frac{\mathbf{x}^T\boldsymbol{\Delta}^{-1}\mathbf{x}}{2}\right). \tag{4.17}$$

The algorithm for generating samples asymptotically distributed according to the posterior distribution using Gibbs sampler is shown in Algorithm 4.

Algorithm 4 Deconvolution via MCMC: Gibbs Sampling Algorithm

1: **Fixed input parameters**: Number of burn-in iterations N_{bi}, total number of iterations N_{MC}
2: **Initializations** $(k = 0)$
- Set $\mathbf{x}^{(0)}$, $\sigma^{2(0)}$, $\beta^{(0)}$, $\gamma^{2(0)}$

3: **Repeat** $(1 \leq k \leq N_{\text{MC}})$
- Sample $\mathbf{x}^{(k)}|(\mathbf{y}, \sigma^{2(k-1)}, \gamma^{2(k-1)})$ from Eq.(4.13)
- Sample $\sigma^{2(k)}|(\mathbf{y}, \mathbf{x}^{(k)}, \beta^{(k-1)})$ from Eq.(4.15)
- Sample $\beta^{(k)}|\sigma^{2(k)}$ from Eq.(4.16)
- Sample $\gamma^{2(k)}|\mathbf{x}^{(k)}$ from Eq.(4.17)

4: **Set** $k = k + 1$.

The posterior distribution mean or minimum mean square error (MMSE) estimator of $\mathbf{x}$ can be approximated by

$$\hat{\mathbf{x}} = \frac{1}{N_{\text{MC}} - N_{\text{bi}}} \sum_{t=N_{\text{bi}}+1}^{N_{\text{MC}}} \mathbf{x}^{(t)}, \tag{4.18}$$

where the samples from the first N_{bi} iterations (corresponding to the transient regime or burn-in

period, which is determined visually from preliminary runs) of the sampler are discarded.

4.4.2 Variational Bayes algorithm

For this approach, we consider an approximation of $p(\boldsymbol{\Theta}|\mathbf{y})$ by a simpler tractable distribution $q(\boldsymbol{\Theta})$ following the variational methodology [23]. Moreover, here, we relax the positivity constraints about the intensity field vector $\mathbf{x}$. Note, however that the positivity constraints can be incorporated but the covariance matrix of the intensity field $\mathbf{x}$ would become more complicated [172, chap. 5]. As will be shown in Sections 4.6 and 4.7, this constraint relaxation yields a fast estimation procedure providing estimation results which compete with the methods incorporating this constraint. The distribution $q(\boldsymbol{\Theta})$ will be found by minimizing the Kullback-Leibler (KL) divergence between the actual posterior distribution and its approximation, given by [99] [100]

$$\mathbf{D}_{\mathrm{KL}}\left(q(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\mathbf{y})\right) = \int q(\boldsymbol{\Theta}) \log\left(\frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\mathbf{y})}\right) d\boldsymbol{\Theta}, \tag{4.19}$$

which is always non-negative and equal to zero only when $q(\boldsymbol{\Theta}) = p(\boldsymbol{\Theta}|\mathbf{y})$. In order to obtain a tractable approximation, the family of distributions $q(\boldsymbol{\Theta})$ are restricted utilizing the mean field approximation [145] so that $q(\boldsymbol{\Theta}) = q(\boldsymbol{\phi})q(\mathbf{x})q(\sigma^2)$, where $q(\boldsymbol{\phi}) = q(\gamma^2)q(\beta)$.

The lower bound of the KL divergence is given by

$$p(\boldsymbol{\Theta}, \mathbf{y}) \geq p(\mathbf{y}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\boldsymbol{\phi})p(\boldsymbol{\phi}) = F(\boldsymbol{\Theta}, \mathbf{y}). \tag{4.20}$$

For $\mathcal{H} \in \{x, \sigma^2, \gamma^2, \beta\}$, let us denote by $\boldsymbol{\Theta}_{\setminus\mathcal{H}}$, the subset of $\boldsymbol{\Theta}$ with $\mathcal{H}$ removed; for instance, if $\mathcal{H} = \mathbf{x}$, $\boldsymbol{\Theta}_{\setminus\mathbf{x}} = \{\sigma^2, \gamma^2, \beta\}$. Then utilizing the lower bound $\mathbf{F}(\boldsymbol{\Theta}, \mathbf{y})$ for the joint probability distribution in Eq.(4.19) we obtain an upper bound for the KL divergence as follows

$$\begin{aligned}
\mathcal{M}\left(q(\boldsymbol{\Theta})\right) &= \int q(\boldsymbol{\Theta}) \log\left(\frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\mathbf{y})}\right) d\boldsymbol{\Theta} \\
&\leq \int q(\mathcal{H}) \left(\int q(\boldsymbol{\Theta}_{\setminus\mathcal{H}}) \log\left(\frac{q(\mathcal{H})q(\boldsymbol{\Theta}_{\setminus\mathcal{H}})}{F(\boldsymbol{\Theta}, \mathbf{y})}\right) d\boldsymbol{\Theta}_{\setminus\mathcal{H}} \right) d\mathcal{H} \\
&= \mathcal{M}\left(q(\mathcal{H})\right).
\end{aligned} \tag{4.21}$$

Therefore, we minimize this upper bound instead of minimizing the KL divergence in Eq.(4.19). Note that the form of the inequality in Eq.(4.21) suggests an alternating (cyclic) optimization strategy where the algorithm cycles through the unknown distributions and replaces each variable with a revised estimate given by the minimum of Eq.(4.21) with the other distributions held constant. Thus, given $q(\boldsymbol{\Theta}_{\setminus\mathcal{H}})$, the posterior distribution approximation $q(\mathcal{H})$ can be computed by solving

$$\hat{q}(\mathcal{H}) = \underset{q(\mathcal{H})}{\text{minimize}} \quad \mathbf{D}_{\mathrm{KL}}\left(q(\boldsymbol{\Theta}_{\setminus\mathcal{H}})q(\mathcal{H})||F(\boldsymbol{\Theta}, \mathbf{y})\right). \tag{4.22}$$

In order to solve this equation, we note that differentiating the integral on the right hand side in Eq.(4.21) w.r.t. $q(\mathcal{H})$ results in (see [127], Eq. (2.28))

$$\hat{q}(\mathcal{H}) = \text{const.} \times \exp\left(E_{q(\boldsymbol{\Theta}_{\setminus\mathcal{H}})}[\log F(\boldsymbol{\Theta}, \mathbf{y})]\right), \tag{4.23}$$

where

$$Eq(\boldsymbol{\Theta}_{\setminus\mathcal{H}})[\log F(\boldsymbol{\Theta}, \mathbf{y})] = \int \log F(\boldsymbol{\Theta}, \mathbf{y}) q(\boldsymbol{\Theta}_{\setminus\mathcal{H}}) d\boldsymbol{\Theta}_{\setminus\mathcal{H}}. \tag{4.24}$$

We obtain the following iterative procedure to find $q(\boldsymbol{\Theta})$ by applying this minimization to each unknown in an alternating way.

Algorithm 5 VB algorithm

1: **Set** $k = 1$, **choose** $q^1(\sigma^2), q^1(\beta)$ and $q^1(\gamma^2)$, initial estimates of the distributions $q(\sigma^2), q(\beta)$ and $q(\gamma^2)$,
2: **repeat** (k = k + 1)
3: $q^k(\mathbf{x}) = \underset{q(\mathbf{x})}{\text{minimize}} \int\int q^k(\boldsymbol{\Theta}_{\setminus\mathbf{x}}) q(\mathbf{x}) \times \log\left(\frac{q^k(\boldsymbol{\Theta}_{\setminus\mathbf{x}}) q(\mathbf{x})}{F(\boldsymbol{\Theta}^k_{\setminus\mathbf{x}}, \mathbf{x}, \mathbf{y})}\right) d\boldsymbol{\Theta}_{\setminus\mathbf{x}} d\mathbf{x}$
4: $q^k(\sigma^2) = \underset{q(\sigma^2)}{\text{minimize}} \int\int q^k(\boldsymbol{\Theta}_{\setminus\sigma^2}) q(\sigma^2) \times \log\left(\frac{q^k(\boldsymbol{\Theta}_{\setminus\sigma^2}) q(\sigma^2)}{F(\boldsymbol{\Theta}^k_{\setminus\sigma^2}, \mathbf{x}, \mathbf{y})}\right) d\boldsymbol{\Theta}_{\setminus\sigma^2} d\sigma^2$
5: $q^k(\gamma^2) = \underset{q(\gamma^2)}{\text{minimize}} \int\int q^k(\boldsymbol{\Theta}_{\setminus\gamma^2}) q(\gamma^2) \times \log\left(\frac{q^k(\boldsymbol{\Theta}_{\setminus\gamma^2}) q(\gamma^2)}{F(\boldsymbol{\Theta}^k_{\setminus\gamma^2}, \mathbf{x}, \mathbf{y})}\right) d\boldsymbol{\Theta}_{\setminus\gamma^2} d\gamma^2$
6: $q^k(\beta) = \underset{q(\beta)}{\text{minimize}} \int\int q^k(\boldsymbol{\Theta}_{\setminus\beta}) q(\beta) \times \log\left(\frac{q^k(\boldsymbol{\Theta}_{\setminus\beta}) q(\beta)}{F(\boldsymbol{\Theta}^k_{\setminus\beta}, \beta, \mathbf{y})}\right) d\boldsymbol{\Theta}_{\setminus\beta} d\beta$
7: **until** some stopping criterion is satisfied.

Now we detail the solutions at each step of algorithm Eq.(5) explicitly.

4.4.2.1 Updating the intensity field vector x

From Eq.(4.23), it can be shown that $q^k(\mathbf{x})$ is an N_1-dimensional Gaussian distribution, rewritten as

$$q^k(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}; E_{q^k(\mathbf{x})}(\mathbf{x}), \boldsymbol{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x})\right), \tag{4.25}$$

where the mean $E_{q^k(\mathbf{x})}(\mathbf{x})$ and covariance $\boldsymbol{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x})$ of this normal distribution can be calculated from step 3 in Algorithm 5 as

$$E_{q^k(\mathbf{x})}(\mathbf{x}) = \frac{(\boldsymbol{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x}))^T \mathbf{H}^T \mathbf{y}}{E_{q^k(\sigma^2)}(\sigma^2)}, \tag{4.26a}$$

$$\boldsymbol{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x}) = \left(\frac{\mathbf{H}^T\mathbf{H}}{E_{q^k(\sigma^2)}(\sigma^2)} + \frac{\boldsymbol{\Delta}^{-1}}{E_{q^k(\gamma^2)}(\gamma^2)}\right)^{-1}. \tag{4.26b}$$

4.4.2.2 Updating the noise variance σ^2

It is easy to show from Eq.(4.23) that the noise variance follows an inverse-Gamma distribution given by

$$q^k(\sigma^2) = \mathcal{IG}\left(\sigma^2; \frac{N_1}{2} + \alpha, E_{q^k(\beta)}(\beta) + E_{q^k(\mathbf{x})}\left[\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right]\right), \tag{4.27}$$

whose mean is given by

$$E_{q^k(\sigma^2)}(\sigma^2) = \frac{E_{q^k(\beta)}(\beta) + E_{q^k(\mathbf{x})}\left[\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right]}{N_1/2 + \alpha - 1}, \tag{4.28}$$

where

$$E_{q^k(\mathbf{x})}\left[\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right] = \|\mathbf{y} - \mathbf{H}E_{q^k(\mathbf{x})}(\mathbf{x})\|_2^2 + \mathrm{tr}\left(\mathbf{H}^T\mathbf{H}\boldsymbol{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x})\right), \tag{4.29}$$

where tr(.) denotes the trace of the matrix.

4.4.2.3 Updating the regularization parameter γ^2

In a similar fashion to the noise variance, the regularization parameter γ^2 follows an inverse-Gamma distribution given by

$$q^k(\gamma^2) = \mathcal{IG}\left(\gamma^2; \frac{N_1}{2} + \eta, \nu + \frac{1}{2}E_{q^k(\mathbf{x})}\left[\mathbf{x}^T\boldsymbol{\Delta}^{-1}\mathbf{x}\right]\right), \tag{4.30}$$

whose mean is given by

$$E_{q^k(\gamma^2)}(\gamma^2) = \frac{\nu + \frac{1}{2}E_{q^k(\mathbf{x})}\left[\mathbf{x}^T\boldsymbol{\Delta}^{-1}\mathbf{x}\right]}{N_1/2 + \eta - 1} \tag{4.31}$$

where

$$E_{q^k(\mathbf{x})}\left[\mathbf{x}^T\boldsymbol{\Delta}^{-1}\mathbf{x}\right] = E_{q^k(\mathbf{x})}(\mathbf{x}^T)\boldsymbol{\Delta}^{-1}E_{q^k(\mathbf{x})}(\mathbf{x}) + \mathrm{tr}\left(\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x})\right). \tag{4.32}$$

4.4.2.4 Updating the hyperparameter β

The hyperparameter β follows a Gamma distribution given by

$$q^k(\beta) = \mathcal{G}\left(\beta; \alpha + \alpha_o, \frac{\beta_o E_{q^k(\sigma^2)}(\sigma^2)}{\beta_o + E_{q^k(\sigma^2)}(\sigma^2)}\right), \tag{4.33}$$

whose mean is given by

$$E_{q^k(\beta)}(\beta) = \frac{(\alpha + \alpha_o)\,\beta_o E_{q^k(\sigma^2)}(\sigma^2)}{\beta_o + E_{q^k(\sigma^2)}(\sigma^2)}. \tag{4.34}$$

In Algorithm 5, no assumptions were imposed on the posterior distribution approximation of $q(\mathbf{x})$. We can, however, assume as [18, 19, 21, 28, 130], that this distribution is degenerate, i.e., distribution which takes one value with probability one and the rest of the values with probability zero. Under this assumption, we can obtain algorithm 6, which is similar to algorithm 5. The

stopping criterion we use is $\sum_{\mathcal{H}\in\{\mathbf{x},\sigma^2,\beta,\gamma^2\}}\|\mathcal{H}^{(k)}-\mathcal{H}^{(k+1)}\|_F \leq \epsilon$, where $\epsilon=\sqrt{N_1}\times 10^{-5}$ [3].

Algorithm 6 Deconvolution via VB

1: **Set** $k=1$,
2: **Initialize** $E_{q^1(\sigma^2)}(\sigma^2), E_{q^1(\gamma^2)}(\gamma^2)$ and $E_{q^1(\beta)}(\beta)$,
3: **repeat** (k = k + 1)
4: $E_{q^k(\mathbf{x})}(\mathbf{x}) = \left(\frac{\mathbf{H}^T\mathbf{H}}{E_{q^k(\sigma^2)}(\sigma^2)} + \frac{\boldsymbol{\Delta}^{-1}}{E_{q^k(\gamma^2)}(\gamma^2)}\right)^{-1}\frac{\mathbf{H}^T\mathbf{y}}{E_{q^k(\sigma^2)}(\sigma^2)}$
5: $E_{q^k(\sigma^2)}(\sigma^2) = \frac{E_{q^k(\beta)}(\beta)+\|\mathbf{y}-\mathbf{H}E_{q^k(\mathbf{x})}(\mathbf{x})\|_2^2}{N_1/2+\alpha-1}$
6: $E_{q^k(\gamma^2)}(\gamma^2) = \frac{\nu+\frac{1}{2}\left(E_{q^k(\mathbf{x})}(\mathbf{x})\right)^T\boldsymbol{\Delta}^{-1}E_{q^k(\mathbf{x})}(\mathbf{x})}{N_1/2+\eta-1}$
7: $E_{q^k(\beta)}(\beta) = \frac{(\alpha+\alpha_o)\beta_o E_{q^{k+1}(\sigma^2)}(\sigma^2)}{\beta_o+E_{q^{k+1}(\sigma^2)}(\sigma^2)}$
8: **until** some stopping criterion is satisfied.
9: **Set** $\hat{\mathbf{x}}=E_{q^k(\mathbf{x})}(\mathbf{x})$, $\hat{\sigma^2}=E_{q^k(\sigma^2)}(\sigma^2)$, $\hat{\gamma^2}=E_{q^k(\gamma^2)}(\gamma^2)$, and $\hat{\beta}=E_{q^k(\beta)}(\beta)$

It is clear that using degenerate distribution for $q(\mathbf{x})$ in Algorithm 6 removes the uncertainty terms of the intensity field estimate. However, it has been shown that this helps to improve the restoration performance [18,19,21,28,130]. Moreover, it also reduces the computational complexity as there is no need to compute explicitly the covariance matrix $\boldsymbol{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x})$ at each iteration. Finally, a few remarks are needed to obtain a fast algorithm. The inverse of the covariance matrix $\boldsymbol{\Delta}$ needs to be computed only once before the loop in Algorithm 6. We also considered the MATLAB operation $\left(\frac{\mathbf{H}^T\mathbf{H}}{E^k(\sigma^2)}+\frac{\boldsymbol{\Delta}^{-1}}{E^k(\gamma^2)}\right)\backslash(\mathbf{H}^T\mathbf{y})$ for the update of the intensity field vector $\mathbf{x}$, which is faster than computing the covariance matrix in (4.26b), then updating the mean in (4.26a). For very big images, diagonal approximation [18] or conjugate gradient [116] can be considered for the update of the intensity field vector $\mathbf{x}$.

4.4.3 ADMM algorithm

This section describes another alternative to the MCMC algorithm which is based on a convex optimization algorithm. The latter maximizes the joint posterior distribution Eq.(4.11) $f(\boldsymbol{\Omega}|\mathbf{y},\boldsymbol{\phi})$ with respect to (w.r.t.) the parameters of interest, with fixing the hyperparameter vector $\boldsymbol{\phi}$, to approximate the MAP estimator of $\boldsymbol{\Theta}$, or equivalently, by minimizing the negative log-posterior distribution given by $\mathcal{F}=-\log\left[f(\boldsymbol{\Theta}|\mathbf{y})\right]$. The resulting optimization problem is tackled using ADMM that sequentially updates the different parameters while the others are fixed, which is widely used in the literature for solving imaging inverse problems [2,3,64]. The rationale behind choosing a proximal-based method like ADMM are (i) the non-differentiability of the problem due to the use of non-smooth prior distributions (ii) it is suitable for our problem format, (ii) it has proven convergence properties, and (iv) it leads to a simple, coordinate-descent structure.

The idea of ADMM depends on variable splitting to break down the optimization problem into

smaller ones we can deal with in an iterative way. Consider an unconstrained problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad f_1(\mathbf{x}) + f_2(\mathbf{L}\mathbf{x}), \tag{4.35}$$

where $\mathbf{L}$ is the regularization operator, and f_1 and f_2 are two closed, proper and convex functions. Variable splitting is a simple procedure that consists in creating a new variable, say $\mathbf{u}$, to serve as an argument of f_2, under the constraint that $\mathbf{u} = \mathbf{L}\mathbf{x}$, i.e.,

$$\begin{aligned} \underset{\mathbf{x},\mathbf{u}}{\text{minimize}} \quad & f_1(\mathbf{x}) + f_2(\mathbf{u}), \\ \text{subject to} \quad & \mathbf{u} = \mathbf{L}\mathbf{x}. \end{aligned} \tag{4.36}$$

The rationale behind variable splitting is that it might be easier to solve the constrained problem in Eq.(4.36) than it to solve its equivalent unconstrained counterpart Eq.(4.35). The constrained problem (4.36) is attacked using a quadratic penalty approach, i.e., by solving

$$(\mathbf{x}^{(k+1)}, \mathbf{u}^{(k+1)}) \quad \in \quad \underset{\mathbf{x},\mathbf{u}}{\arg\min}\ f_1(\mathbf{x}) + f_2(\mathbf{u}) + \frac{\mu}{2}\left\|\mathbf{L}\mathbf{x} - \mathbf{u} - \mathbf{d}^{(k)}\right\|_2^2, \tag{4.37}$$

$$\mathbf{d}^{(k+1)} \quad = \mathbf{d}^{(k)} - (\mathbf{L}\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)}), \tag{4.38}$$

where $\mathbf{d}$ is the Lagrange multipliers. The rationale of these methods is that each step of this alternating minimization may be much easier than the original unconstrained problem Eq.(4.35). The ADMM algorithm corresponding to problem Eq.(4.36) (also Eq.(4.35)) is shown in Algorithm 7.

Algorithm 7 ADMM Algorithm

1: set $k = 0$, choose $\mu > 0, \mathbf{u}_0, \mathbf{d}_0$,
2: **repeat** (k = k + 1)
3: $\mathbf{x}^{(k+1)} \in \underset{\mathbf{x}}{\arg\min}\ f_1(\mathbf{x}) + \frac{\mu}{2}\left\|\mathbf{L}\mathbf{x} - \mathbf{u}^{(k)} - \mathbf{d}^{(k)}\right\|_2^2$,
4: $\mathbf{u}^{(k+1)} \in \underset{\mathbf{u}}{\arg\min}\ f_2(\mathbf{u}) + \frac{\mu}{2}\left\|\mathbf{L}\mathbf{x}^{(k+1)} - \mathbf{u} - \mathbf{d}^{(k)}\right\|_2^2$,
5: **Update Lagrange multipliers:**
6: $\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} - (\mathbf{L}\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)})$
7: **Update iteration** $k \leftarrow k + 1$
8: **until** some stopping criterion is satisfied.

where $\mu > 0$ is a constant, which controls convergence speed. The convergence of the ADMM algorithm is guaranteed if $\mathbf{L}$ has full column rank, the functions f_1 and f_2 are closed, proper, and convex functions [3, 56].

Now, we can rewrite our model as an optimization problem as follows

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}) + i_{\mathbb{R}+}(\mathbf{x}), \tag{4.39}$$

where the regularization function $\phi(\mathbf{x})$ is proportional to the negative logarithm of the intensity field prior distribution considered in Eq.(4.6) up to an additive constant, i.e. $\phi(\mathbf{x}) = \frac{\mathbf{x}^T \boldsymbol{\Delta}^{-1} \mathbf{x}}{2}$, and $\lambda = \sigma^2/\gamma^2$ is the regularization parameter. Given this objective function, we write the constrained equivalent formulation as follows

$$\begin{aligned} \underset{\mathbf{u},\mathbf{x}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}) + i_{\mathbb{R}+}(\mathbf{u}), \\ \text{subject to} \quad & \mathbf{u} = \mathbf{x}, \end{aligned} \tag{4.40}$$

where $\mathbf{u}$ and $\mathbf{x}$ are the variables to minimize over. In order to solve for $\mathbf{u}$ and $\mathbf{x}$, we construct the augmented Lagrangian corresponding to Eq.(4.40) as follows

$$\mathcal{L}(\mathbf{u}, \mathbf{x}, \mathbf{d}_1) = \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}) + i_{\mathbb{R}+}(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{u} - \mathbf{d}_1\|_2^2. \tag{4.41}$$

The ADMM algorithm for solving Eq.(4.41) is shown in Algorithm 8. During each step of the iterative algorithm, $\mathcal{L}$ is optimized w.r.t. $\mathbf{u}$ (step 3) and $\mathbf{x}$ (step 4) and then the Lagrange multipliers are updated (step 6). From preliminary runs, we fixed μ to $\mu = 10$ which leaded to fast convergence, it can however be updated within the algorithm as in [32]. The stopping criterion we use is $\|\mathbf{u}^{(k)} - \mathbf{x}^{(k)}\|_2 \leq \epsilon$, where $\epsilon = \sqrt{N_1} \times 10^{-5}$ [3].

Algorithm 8 Deconvolution via ADMM

1: set $k = 0$, choose $\mu > 0, \mathbf{u}^{(0)}, \mathbf{x}^{(0)}$, and $\mathbf{d}_1^{(0)}$
2: **repeat** (k = k + 1)
3: $\mathbf{u}^{(k+1)} = \max\left(\mathbf{x}^{(k)} - \mathbf{d}_1^{(k)}, 0\right)$
4: $\mathbf{x}^{(k+1)} = \left(\mathbf{H}^T\mathbf{H} + \lambda\boldsymbol{\Delta}^{-1} + \mu\mathbf{I}\right)^{-1}\left[\mathbf{H}^T\mathbf{y} + \mu\left(\mathbf{u} + \mathbf{d}_1^{(k)}\right)\right]$
5: **Update Lagrange multipliers:**
6: $\mathbf{d}_1^{(k+1)} = \mathbf{d}_1^{(k)} - \left(\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)}\right)$
7: **Update iteration** $k \leftarrow k + 1$
8: **until** some stopping criterion is satisfied.

4.5 Non-Linear Interpolation Using Gaussian Process Regression

In order to visually view a meaningful image from the deconvolved intensities, we consider non-linear interpolation based on Gaussian processes (GP) [150], since it can provide confidence intervals for each interpolated pixel. A classic choice consists of considering a zero-mean GP with an arbitrary covariance matrix. Here, we choose this covariance matrix to be $\boldsymbol{\Delta}' = \boldsymbol{\Delta}/\gamma^2$. Precisely, we interpolate using the prior distribution previously defined in Eq.(4.7). If $d_{n,n'}$ is very small, then $\boldsymbol{\Delta}'(n, n')$ approaches its maximum $1/\gamma^2$. If n is distant from n', we have instead $\boldsymbol{\Delta}'(n, n') \approx 0$, i.e. the two points are considered to be *a priori* independent. For example, during interpolation

at new n_* location, distant cores will have negligible effect. The amount of spatial correlation depends on the parameters ℓ, and κ, which are estimated in the way we previously mentioned in Section 4.3.2.1.

If we consider $\mathbf{\Delta}'(\mathbf{z},\mathbf{z}) \in \mathbb{R}^{N_1 \times N_1}$, $\mathbf{z} = [z_1, \ldots, z_{N_1}]^T$ contains all the positions of all the observed cores (whose estimated intensities are gathered into $\mathbf{x}$), and a new spatial location z_* for which we want to predict the intensity x_*, the GP can be extended as follows

$$\begin{bmatrix} \mathbf{x} \\ x_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{\Delta}'(\mathbf{z},\mathbf{z}) & \mathbf{\Delta}'(\mathbf{z},z_*) \\ \mathbf{\Delta}'(z_*,\mathbf{z}) & 1/\gamma^2 \end{bmatrix}\right), \tag{4.42}$$

where $\mathbf{\Delta}'(\mathbf{z},z_*) = \mathbf{\Delta}'(z_*,\mathbf{z})^T \in \mathbb{R}^{N_1}$. Eq.(4.42) shows that the conditional distribution of each predicted intensity given the previously estimated intensities, follows a Gaussian distribution $x_*|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ whose mean and variance are given by

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{\Delta}'(z_*,\mathbf{z})\mathbf{\Delta}'(\mathbf{z},\mathbf{z})^{-1}\mathbf{x}, \\ \mathbf{\Sigma} &= 1/\gamma^2 - \mathbf{\Delta}'(z_*,\mathbf{z})\mathbf{\Delta}'(\mathbf{z},\mathbf{z})^{-1}\mathbf{\Delta}'(\mathbf{z},z_*). \end{aligned} \tag{4.43}$$

By setting $\mathbf{x} = \hat{\mathbf{x}}$, the mean in Eq.(4.43) is finally used to estimate each interpolated intensity, while the variance is used to provide additional information (measure of uncertainty) about the interpolated intensity values. Note that the uncertainty of the deconvolved vector $\hat{\mathbf{x}}$ computed in the MCMC and VB cases can be incorporated in the interpolation process. However, the variance vector $\mathbf{\Delta}'(\mathbf{z},z_*)$ of a new sample will require taking into account the cross coupling effect between this new sample and the deconvolved fibre core intensities which does not exist. Moreover, the uncertainty computed in MCMC will require high number of samples to the deconvolved vector $\hat{\mathbf{x}}$ in order to get a good approximation.

4.6 Simulations Using Synthetic Data

4.6.1 Data creation

The performance of the proposed methods is investigated by reconstructing a standard test image. A subsampled version of this image is obtained by considering the sampling pattern of an actual endomicroscopy system, as illustrated in Fig. 4.7. This figure provides an example of a homogeneous region imaged through Alveoflex (Mauna Kea Technologies, Paris, France) fibre bundle [105] [16]. Such image is used for calibration and to identify the number and positions of the fibre cores. The build-in MATLAB function "vision.BlobAnalysis" was used to detect central fibre core pixels.

Fig. 4.8 shows the original Lena image (left) and an example of system output (right) after applying the model in Eq. Eq.(4.2). This image is formed by creating a binary mask in which a value of 1 is assigned to pixels corresponding to the central pixels of each core in Fig. 4.7(b), and

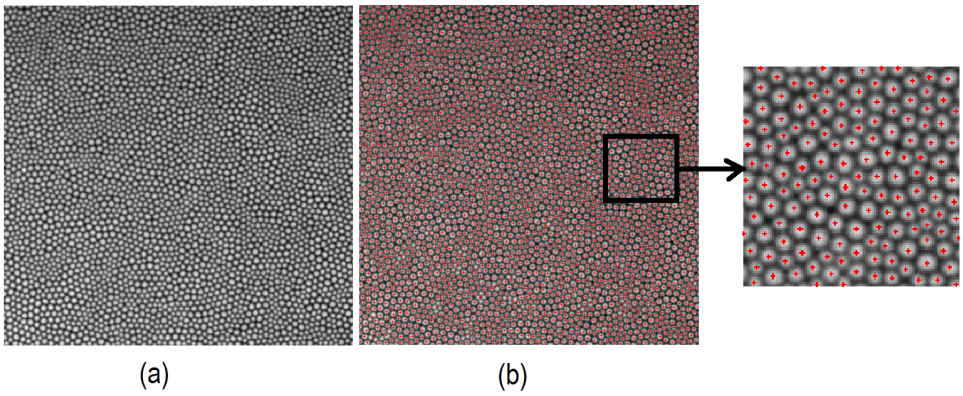


Figure 4.7 (a) Example of 512×512 pixels image of the endomicroscopy system (b) Image with detected fibre core centres superimposed (red crosses).

zero otherwise. This mask is then multiplied point by point by the Lena image in Fig. 4.8(a) in order to obtain the subsampled image. The model in Eq.(4.2) is then applied to obtain an image that simulates the system's output which is shown in Fig. 4.8(b). This image is created using subsampled intensities corresponding to 1.29% of the original Lena image. For simulated data, we considered a Gaussian spatial blurring kernel with one size $\sigma^2_{\mathbf{C}} = 2$ in all the simulations.

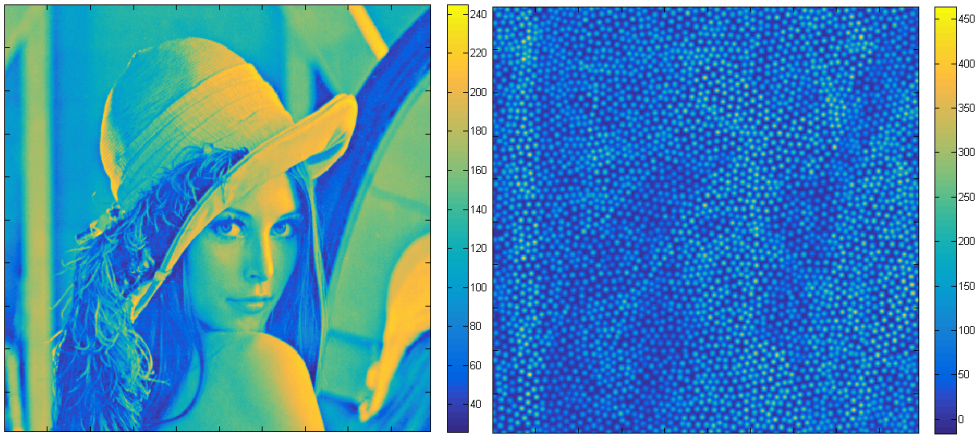


Figure 4.8 Creation of the synthetic data: (a) Original image (b) example of final system output with $\sigma^2_{\mathbf{H}} = 20$ and $\sigma^2 = 10$.

4.6.2 Performance analysis

The performance discriminator adopted in this work to measure the quality of the deconvolved fibre cores is the root mean square error (RMSE), which is computed using intensities at the core locations using

$$\text{RMSE}(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\frac{\sum_{n=1}^{N_1} \left(\mathbf{x}(n) - \hat{\mathbf{x}}(n)\right)^2}{N_1}}, \tag{4.44}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are vectors of the subsampled reference Lena image and its deconvolved version respectively, and N_1 is the number of fibre cores.

For synthetic data, in order to check the performance of the algorithm with different cross coupling effects, different values of $\alpha_{\mathbf{H}}$ and $\beta_{\mathbf{H}}$ in Eq.(4.3) can be considered. However, this can be simplified by considering a 2D Gaussian kernel defined by

$$[\mathbf{H}]_{i,j} = \exp\left(\frac{-d_{i,j}^2}{2\sigma_{\mathbf{H}}^2}\right), \tag{4.45}$$

since it involves only one variable to change, namely $\sigma_{\mathbf{H}}^2$ (representing a squared distance, in pixels). This is equivalent to setting $\beta_{\mathbf{H}} = 2$ and $\alpha_{\mathbf{H}}^2 = \alpha_{\mathbf{H}}^2/2$. Note that this simplification is considered only for synthetic data in order to assess the influence of the kernel width. The generalized Gaussian cross coupling kernel $\mathbf{H}$ defined in Eq.(4.3) will be considered for real data.

The three methods showed similar results in terms of RMSE and interpolated images. The following shows the VB method's results. Fig. 4.9 shows examples of interpolated intensities after deconvolution using GP in the noise-free case ($\sigma^2 = 0$) and noisy case ($\sigma^2 = 10$) and different values of $\sigma_{\mathbf{H}}^2$, with the corresponding confidence interval images. We can observe that the structure of the Lena image can be recovered in the two cases. Moreover, in the confidence interval images, we can observe that as we go away from central cores, the confidence interval of the interpolated intensities decreases.

In order to measure the performance of the algorithms, we consider different noise variances (σ^2) as well as different cross coupling effects ($\sigma_{\mathbf{H}}^2$). Fig. 4.10 shows the RMSE (in log-scale) before and after deconvolution versus $\sigma_{\mathbf{H}}^2$ at $\sigma^2 = 10$. We can observe that all of the methods are very effective since the RMSE after deconvolution is always lower than that before deconvolution. Moreover, the gain increases with cross coupling.

In order to analyze the effect of noise variance and cross coupling separately, we fix one of them and change the other as shown in Fig. 4.11. In this figure, we show plots of RMSEs after deconvolution for different σ^2 at fixed $\sigma_{\mathbf{H}}^2$ and vice versa. In Fig. 4.11(a), we can observe that there is roughly a linear relationship between RMSE and σ^2 at fixed $\sigma_{\mathbf{H}}^2$. Moreover, the behaviour at $\sigma_{\mathbf{H}}^2 = 1, 5, 10$ and 15 is almost the same. In Fig. 4.11(b), we can observe that RMSE is fairly constant as $\sigma_{\mathbf{H}}^2$ increases at constant σ^2. Furthermore, it starts to increase as σ^2 increases but still remains constant when changing $\sigma_{\mathbf{H}}^2$.

For the MCMC method, in all of the simulations in this work including the real datasets, $N_{\text{MC}} = 1500$, including $N_{\text{bi}} = 500$, which were determined visually from preliminary runs, were used. For the ADMM method, different regularization parameter values are tested, we pick up the one corresponding to the lowest RMSE.

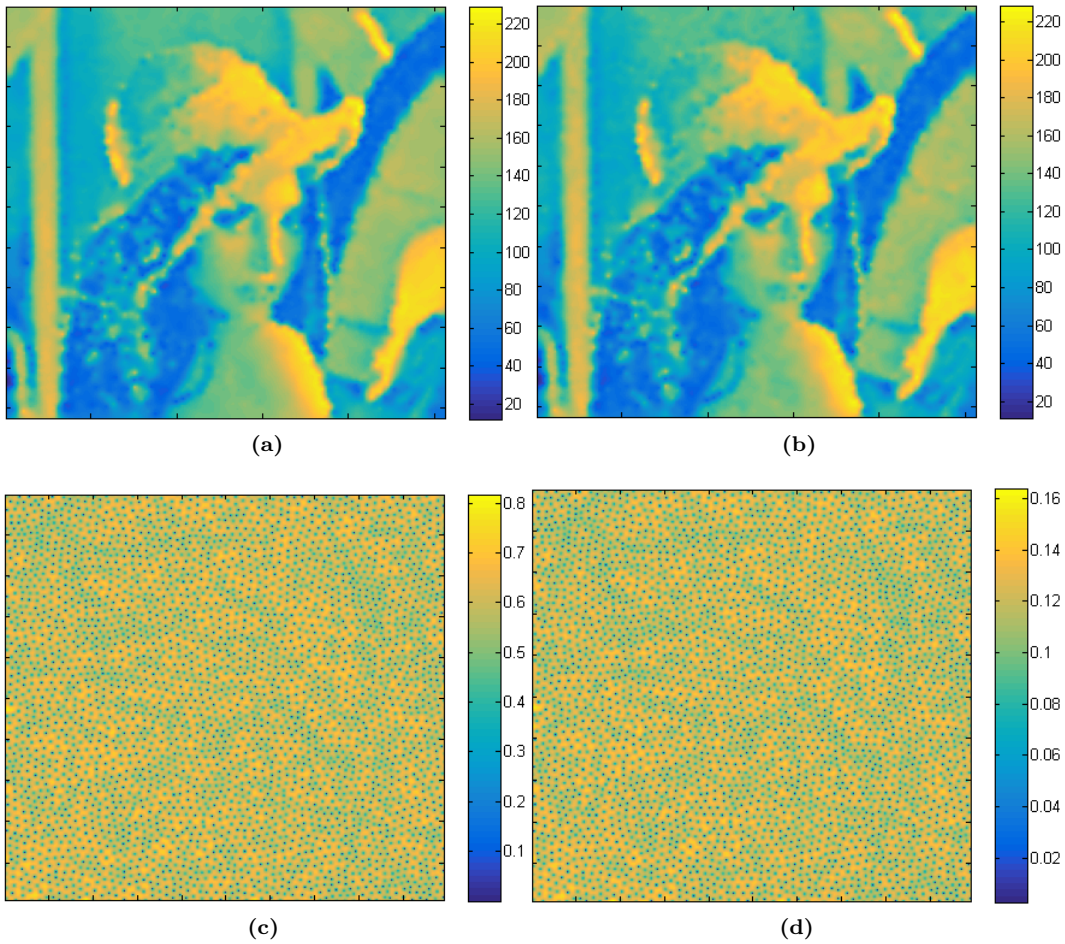


Figure 4.9 Examples of interpolated samples by GP after deconvolution (a) $\sigma^2 = 0$ and $\sigma^2_{\mathbf{H}} = 1$, and (b) $\sigma^2 = 10$ and $\sigma^2_{\mathbf{H}} = 20$, and the corresponding confidence interval images.

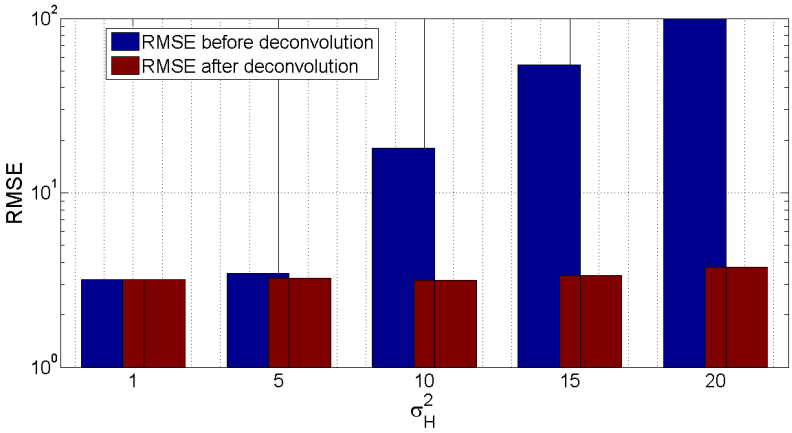


Figure 4.10 Plot of RMSEs before and after deconvolution (in-log scale) versus $\sigma^2_{\mathbf{H}}$ at $\sigma^2 = 10$.

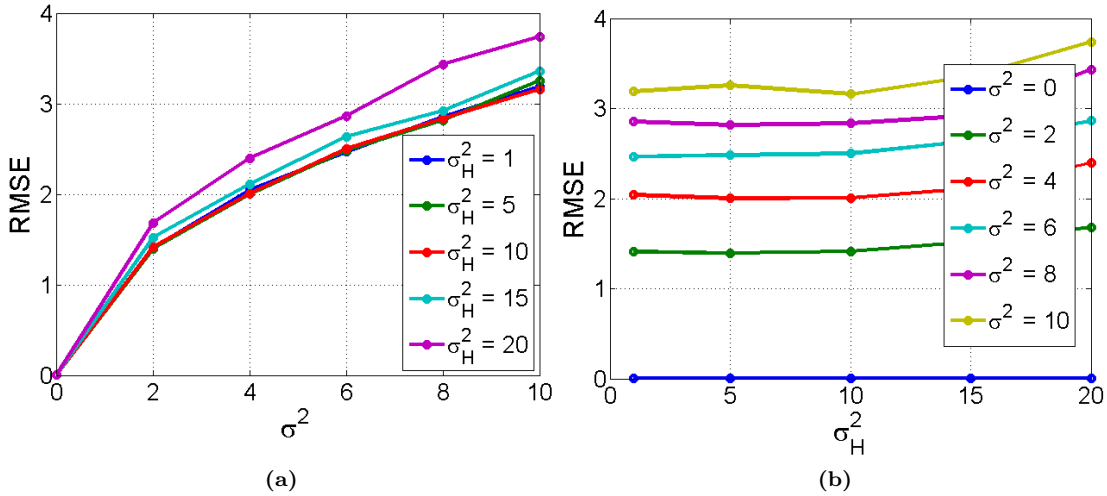


Figure 4.11 Plot of RMSEs after deconvolution (a) versus σ^2 at fixed $\sigma^2_{\mathbf{H}}$, and (b) versus $\sigma^2_{\mathbf{H}}$ at fixed σ^2.

4.6.3 Comparison

In this section, we compare the three proposed methods for deconvolution and restoration of OEM images. The comparison is conducted in terms of RMSE before and after deconvolution, as well as in terms of computation time.

Fig. 4.12 compares RMSEs after deconvolution versus different σ^2 as well as different $\sigma^2_{\mathbf{H}}$. We can observe that for all of the methods, as σ^2 increases at constant $\sigma^2_{\mathbf{H}}$, RMSE increases. On the other hand, at fixed σ^2, RMSE seems to be roughly constant for $\sigma^2_{\mathbf{H}} = 1, 5,$ and 10, then, it starts to increase as $\sigma^2_{\mathbf{H}}$ increases. It is clear that all the methods behave similarly in terms of RMSE.

The computation performance of the algorithms is measured in terms of the running time of each of them, which gives a rough estimation of the actual computation performance. Note, however that floating point operations per second (FLOPS) can also be computed. Table. 4.1 shows the average computation time (in seconds) of the three proposed methods. All the experiments presented in this thesis were conducted on ACER core-i3-2.0 GHz processor laptop with 8 GB RAM. It is clear that the MCMC method is the most computationally expensive method. The ADMM method is second, and the VB the least. Despite the relatively high computation time of the MCMC method, it is a parameter free method compared to the ADMM-based method in which the regularization parameter λ should be chosen carefully. The VB approach is considered to be the best compared to MCMC and ADMM, it can provide similar RMSE but with lower computation complexity, moreover, it is fully automatic in the sense that it can estimate the hyperparameters associated with the parameters as mentioned previously in section 4.4.2.

Although the MCMC and ADMM algorithms can estimate the noise variance and model hyperparameters, in practice these parameters are very difficult to estimate accurately, (specifically σ^2 and γ^2) due to the similarity between $\mathbf{H}^T\mathbf{H}$ and $\mathbf{\Delta}^{-1}$ in (4.14b) and (4.26b). Therefore, we

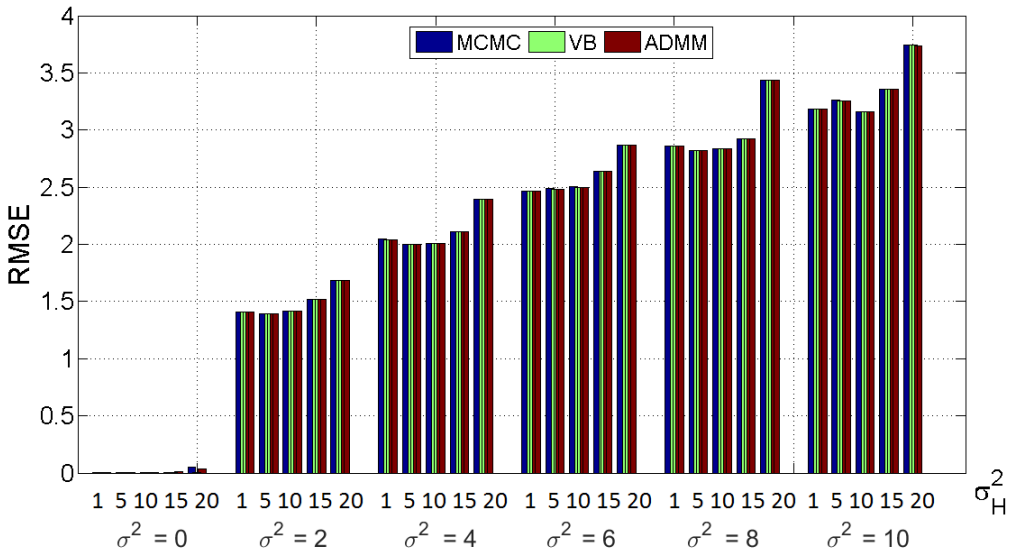


Figure 4.12 Plot of RMSEs before and after deconvolution for the three methods versus σ^2 as well as σ^2_H.

Method	MCMC	ADMM	VB
Computation time (sec.)	3100	35.51	**5.12**

Table 4.1 The average computation time (in seconds) of the three proposed methods. In order to maintain a fair comparison between the three algorithms, the computational time of the ADMM algorithm corresponds to the duration of five runs (used to select the best regularization parameter among the five values).

have to make an informed choice about one of these parameters, specifically the choice of the hyperparameters α, α_0 and β_0 in Eq.(4.8) and Eq.(4.9). In Fig. 4.11(b), we observe that the RMSEs in practise are close to the true noise standard deviation, and hence the noise variance can be inferred.

4.6.4 Robustness

To test the robustness of the proposed methods, we create the data using a specific σ^2_H and we deconvolve using different values. Following this strategy, we create the data using $\sigma^2_{\mathbf{H}} = 10$ and we deconvolve using $\sigma^2_{\mathbf{H}} \in \{6, 8, 10, 12, 14\}$. The three estimation approaches showed similar results.

Fig. 4.13 shows plots of RMSE after deconvolution versus σ^2 at fixed $\sigma^2_{\mathbf{H}}$ and vice versa. In Fig. 4.13(a), we can observe that the noise variance has no effect on the deconvolution in the tested interval as RMSE is constant at fixed $\sigma^2_{\mathbf{H}}$. In Fig. 4.13(b), there is an approximately linear relationship between RMSE and $\sigma^2_{\mathbf{H}}$ at constant σ^2. Furthermore, lower values of $\sigma^2_{\mathbf{H}}$ than the one we created the data with (i.e., $\sigma^2_{\mathbf{H}} = 6$ and 8) yield lower RMSE than higher ones (i.e., $\sigma^2_{\mathbf{H}} = 12$ and 14). In other words, it is slightly better to underestimate $\sigma^2_{\mathbf{H}}$ than to overestimate it.

We observe that deconvolution using the value we created the data with ($\sigma^2_{\mathbf{H}} = 10$) yields the

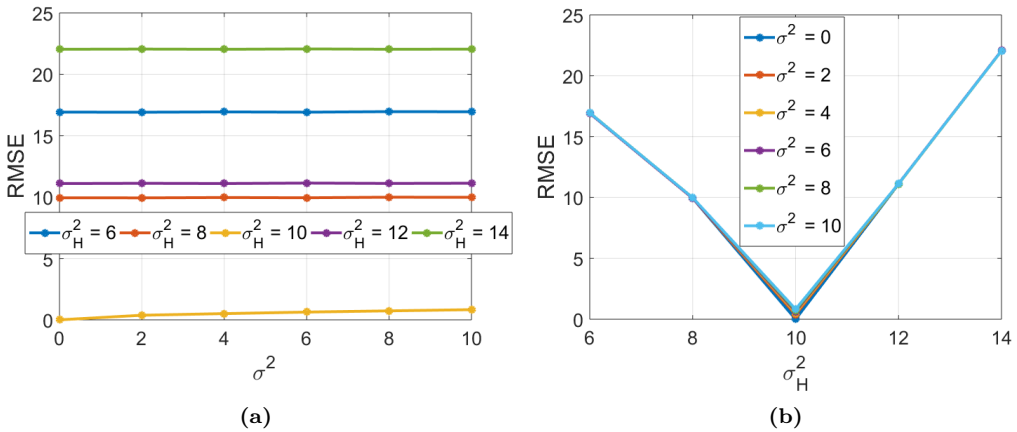


Figure 4.13 Plots of RMSEs between the central fibre cores in the original Lena image and the deconvolved central fibre cores versus (a) σ^2 at fixed σ^2_H, and (b) σ^2_H at fixed σ^2 of the IIo method.

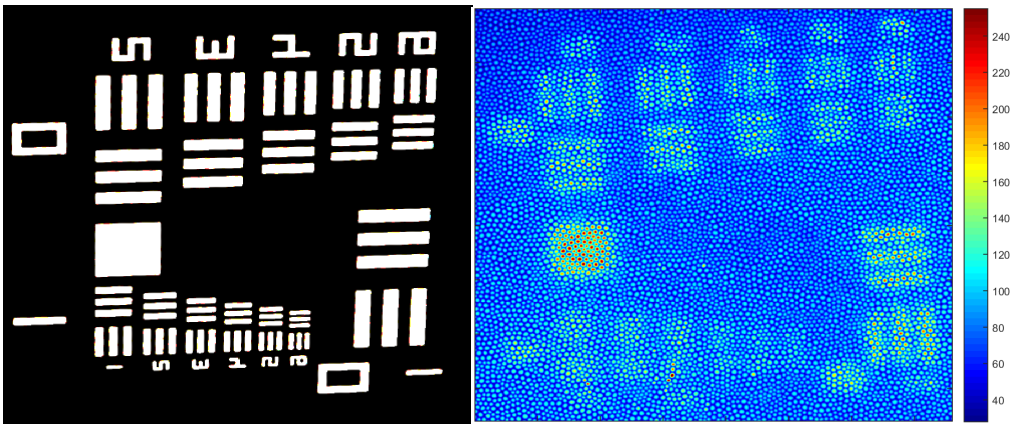
minimum RMSE. Moreover, RMSE after deconvolution is always lower than that before deconvolution except for $\sigma^2_{\mathbf{H}} = 14$ at which it is higher.

4.7 Simulations Using Real Data

The performance of the proposed methods has been evaluated on two real datasets; the 1951 USAF resolution test chart and *ex vivo* human lung tissue. Both of them were collected using OEM system [98] with monochrome detection (Grasshopper3 camera GS3-U3-23S6M-C, Point Grey Research, Canada) and 470 nm LED illumination (M470L3, Thorlabs Ltd, UK) for lung autofluorescence excitation. Excised human lung tissue was placed in a well plate. Human tissue was used with regional ethics committee (REC: 13/ES/0126) approval and was retrieved from the periphery of specimens taken from lung cancer resections. In order to adjust the cross coupling kernel parameters $\alpha_{\mathbf{H}}$ and $\beta_{\mathbf{H}}$, a study was performed to measure, analyze and quantify inter-core coupling within coherent fibre bundles [146]. This study showed how light is spread over the neighbouring cores, and gave statistical analysis on coupling percent in neighbouring cores. It showed that around 61% of transmitted light remains in the central core, around 34% in the first neighbouring cores, around 4% in the second neighbouring cores, and less than 1% in the third, fourth and fifth neighbouring cores. This leads to fixing $\alpha_{\mathbf{H}} = 4$ (in pixels) and $\beta_{\mathbf{H}} = 0.8$.

4.7.1 1951 USAF resolution test chart

The 1951 USAF chart is a resolution test pattern set by US Air Force in 1951. It is widely accepted to test the resolution of optical imaging systems such as microscopes, cameras and image scanners [167]. Fig. 4.14 (a) shows the original USAF resolution test chart used in the project. The resulting image obtained by fibre bundle is shown in Fig. 4.14 (b) with image size 760×760



(a) (b)

Figure 4.14 (a) Scanned image of an USAF 1951 Resolution test chart. (b) The 1951 USAF resolution test chart imaged by the OEM system.

and is composed of 7,776 fibre cores (1.34% of the image).

A non-linear interpolation based on GP of central core intensities of the image in Fig. 4.14(b) is presented in Fig.4.15(a), with the corresponding confidence intervals image in Fig.4.15(c). We can observe the blurring which is caused by the cross coupling effect as well as the sparsity of the data.

The outputs of the MCMC, VB, and ADMM algorithms are very similar. Thus, we show the results of the VB method. Fig. 4.15(b) shows an example of one of the output images with the corresponding confidence intervals in Fig. 4.15(d). The set of thicker strips (top left corner of the image) is now better resolved and the overlap between them is reduced. The small set of strips which is at the bottom could not be resolved, which gives an indication about the resolving resolution of this endomicroscopy system. Regions of high uncertainty (which appear as blobs in dark red) are where there may be no cores or they are dead, this in addition to the irregular core sampling are the reasons for some strips appear a bit fragmented.

4.7.2 *Ex vivo* human lung tissues

Fig. 4.16(a) shows the output image of the OEM system. Image size is 1000×800 and is composed of 13,343 fibre cores (1.66% of the image). Non-linear interpolation based on GP of central core intensities is presented in Fig.4.16(b). Similar to the USAF resolution test chart, we aim at reducing cross coupling effect as well as getting a more resolved image.

Similar to the USAF resolution test chart results, the outputs of the MCMC, VB, and ADMM algorithms are very similar. We only show the results of the VB method. Fig. 4.16(c) shows an example of interpolated deconvolved samples using GP. The lung structure is now better resolved and more sharper than before deconvolution. Moreover, confidence intervals are shown in Fig. 4.16(d). We can observe that as we move away from the central cores, the confidence of the

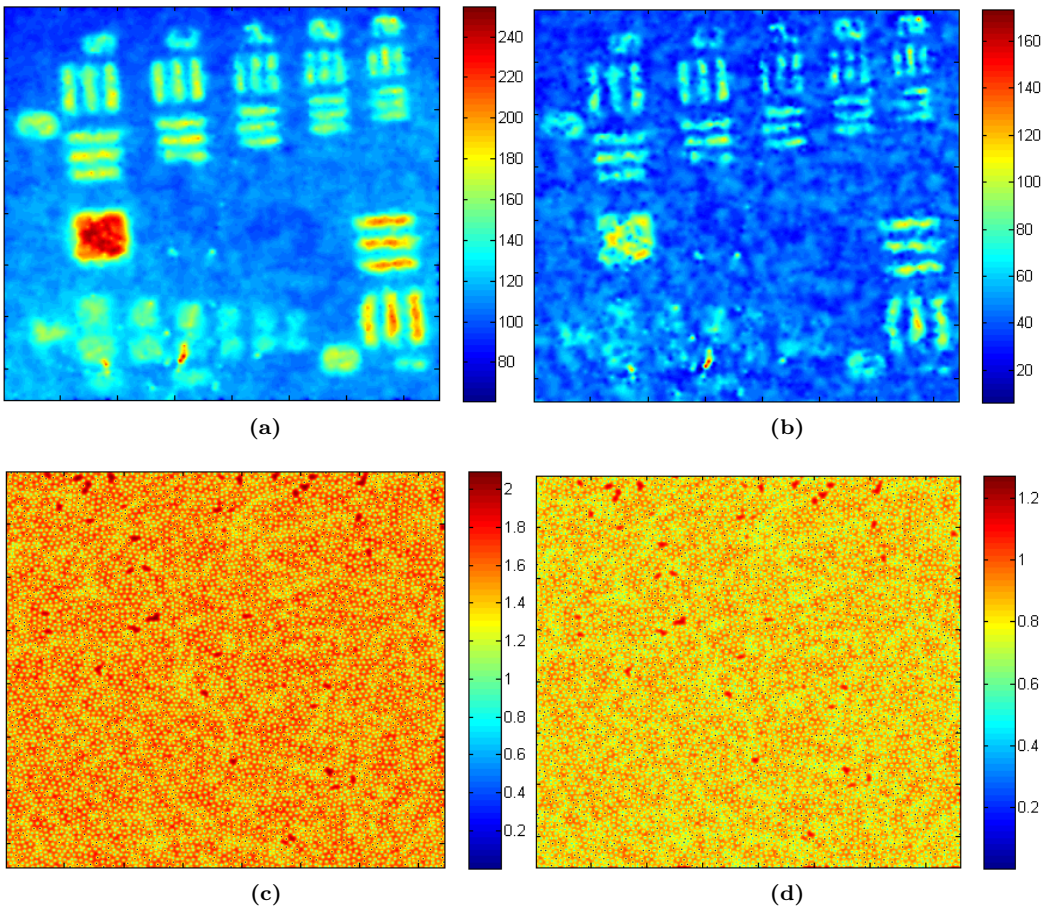


Figure 4.15 Non-linear interpolation (a) before, and (b) after deconvolution, and their corresponding confidence intervals in (c), and (d) respectively.

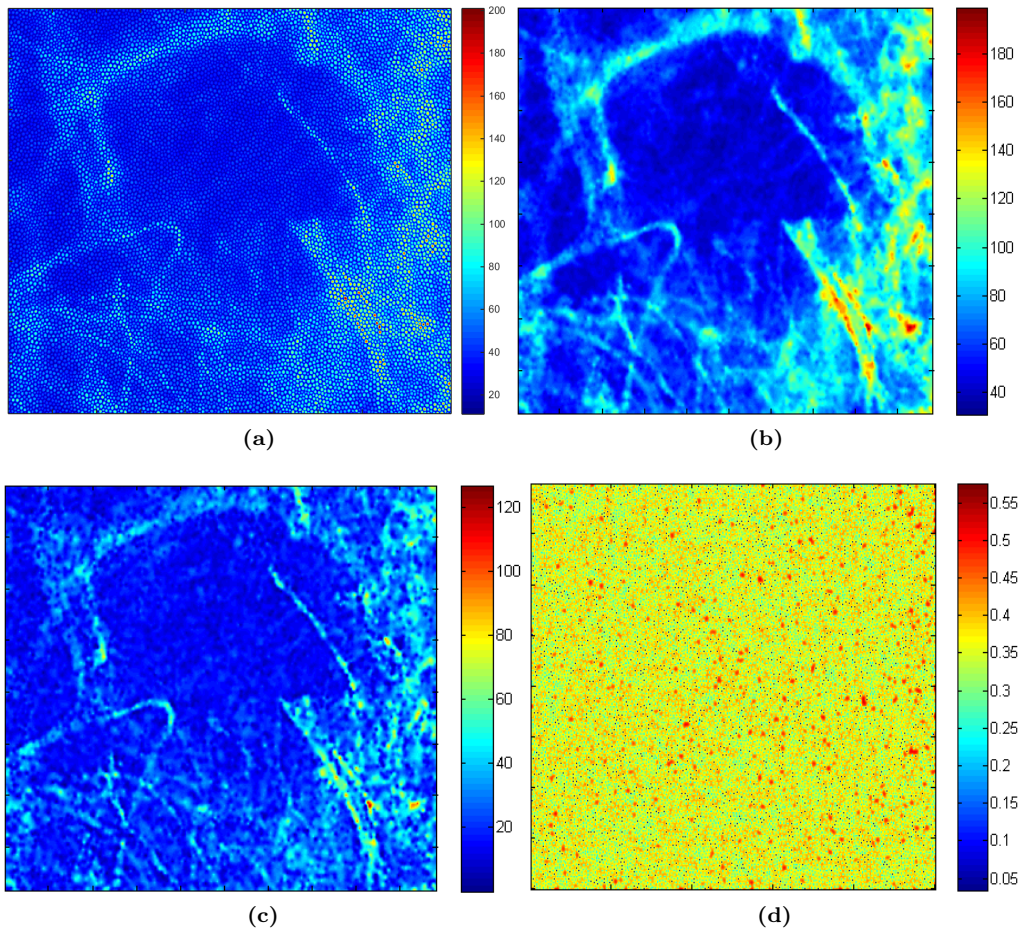


Figure 4.16 (a) *Ex vivo* lung tissue imaged by the endomicroscopy system [98]. Non-linear interpolation (b) before, and (c) after deconvolution, (d) the confidence intervals of the image in (c).

Dataset/Method	MCMC	ADMM	VB
USAF chart	1.12×10^5	250	**5.9**
Lung tissue	1.46×10^6	870	**16.05**

Table 4.2 Computation time (in seconds) for the real data. In order to keep a fair comparison between the three algorithms, the computational times of the ADMM algorithm correspond to the duration of five runs (used to select the best regularization parameter among five values).

interpolated intensities decreases and vice versa.

Table 4.2 provides the computation time of the 1951 USAF resolution test chart and the *ex vivo* lung tissue image. It is clear that the VB is still the fastest despite the change of the images size.

4.8 Conclusions

This chapter introduced a hierarchical Bayesian model and three estimation algorithms for the deconvolution of optical endomicroscopy images. The deconvolution accounts and compensates for fibre core cross coupling which causes image degradation and blurring in this type of imaging. The resulting joint posterior distribution was used to approximate the Bayesian estimators. First, a Markov chain Monte Carlo procedure based on a Gibbs sampler algorithm was used to sample the posterior distribution of interest and to approximate the MMSE estimators of the unknown parameters using the generated samples. Second, a variational Bayes approach to approximate the joint posterior distribution by minimizing the Kullback-Leibler divergence was used. Third, an approach based on an alternating direction method of multipliers was used to approximate the maximum a posteriori estimators. The three algorithms showed similar estimation performance while providing different characteristics, the MCMC and VB based approaches are fully automatic in the sense that they can jointly estimate the hyperparameters associated with the prior distributions, however, the MCMC based approach showed high computational complexity which could be overcome by the VB and ADMM approaches. Although the ADMM approach has low computational complexity, it is semi-supervised in the sense that the hyperparameters associated with the prior distributions need to be chosen carefully by the user. A non-linear interpolation approach based on Gaussian processes was considered to restore the full images from the samples to provide a meaningful image for interpretation. In the future, we can consider temporal information while deconvolving. The VB approach might be a good choice for this online deconvolution approach as it does not only provide faster estimates compared to the other tested methods but also can estimate the model hyperparameters. In the next chapter, we provide a supervised spot-detection approach for the detection of fluorescently labelled bacteria.

Chapter 5

Patch-Based Sparse Representation For Bacterial Detection

This chapter has been adapted from the conference paper [Eldaly et al. (2018d)] (submitted).

Contents

5.1 Introduction

Fibred Confocal Fluorescent optical endomicroscopy, also referred to as probe-based confocal laser endomicroscopy (pCLE), enables the exploration of the distal pulmonary tract as well as the assessment of the respiratory bronchioles and alveolar gas exchanging units of the distal lung [181]. In pulmonary OEM, elastin and collagen generate auto-fluorescence when excited with a 488 nm laser source. While bacteria can potentially auto-fluoresce when exposed to light at the appropriate wavelength(s), the emitted fluorescent signal is usually too weak to be imaged effectively to provide any diagnostic information. To address this limitation, bacteria can be specifically labelled with targeted fluorescently labelled ligands (SmartProbes) [7], which bind to the bacteria to generate fluorescence in response to a light excitation at pre-determined wavelengths (e.g., 488 nm laser excitation) [7]. Stained bacteria appear as bright dots in the images, whereas elastin and collagen are naturally auto-florescent and display a mesh-like structure in the distal lung, making the detection of bacteria a more challenging task.

Bacterial detection and quantification approaches developed for confocal microscopy [109, 191] are not directly applicable to OEM data due to major differences in the image acquisition approaches employed by the two technologies. In particular, confocal microscopy employs a regular rectangular grid to acquire images of high lateral and axial optical resolutions which are 0.18 mm and 0.5 mm respectively [69]. In contrast, OEM employs a sparse and irregular pattern (due to the organisation of the fibres in the fibre bundle) to acquire images of moderate lateral and axial optical resolutions of 0.5 µm and 3 µm respectively within a constantly moving scene (breathing, heart beat and fibre motion) [84].

The bacteria detection problem can be addressed through a variety of different annotation scenarios, two of which are dot-annotation and count-annotation [12, 193]. While dot-annotation provides the location of objects that appear in the image, count-annotation only provides the number of objects in an image without explicitly revealing the locations. Dot-annotations are more informative given the small scale of the objects to be quantified. In this work, we assess the performance of our method by considering the number of detections (count-annotation) but also their estimated locations (dot-annotation).

As previously mentioned in Chapter 3, the object detection problem has been mainly addressed in the literature through either unsupervised or supervised methods. In unsupervised approaches, objects we wish to detect are learned from the data by fitting them with suitable distributions without using explicitly-provided labels. Examples of such methods are hierarchical Bayesian models [9, 53, 124], and linear or morphological filtering methods [33, 81, 96, 170, 171]. On the other hand, in supervised-based approaches [13, 87, 108, 164], the dataset is usually divided into training,

testing and validation sets. In the training phase, a model is trained by pairing inputs with their expected outputs, which are also known as the ground truth. This trained model is then used to estimate the output of the test dataset. The validation set is usually use to tune the model's hyperparameters which can be fixed during the test phase.

In this chapter, we propose an unsupervised-based approach for bacterial detection in OEM images. This approach splits each image into a set of overlapping patches and assumes that the observed pixel fluorescence in each patch is a linear combination of the actual intensity value associated with background image structures, corrupted by additive Gaussian noise and potentially by an additional outlier term modelling anomaly (which is considered to be candidate bacteria). The actual intensity term representing background structures is considered as a weighted linear combination of a few atoms drawn from a dictionary which is learned from bacteria-free data by using a dictionary learning-based approach and then fixed in the rest of the simulations. The real datasets considered in this work have been acquired by the Cellvizio, Mauna Kea Technologies (Paris, France) confocal endomicroscopy system [16], in which a set of preprocessing steps have been applied. This makes the observation noise difficult to model analytically, thus we assume here that the noise is additive, white and Gaussian [124]. The bacteria detection task is formulated by a minimization problem in which suitable regularization functions are assigned to the unknown model parameters. Although the resulting optimization problem is convex, it is not smooth, hence a proximal-based method is required. In this work, we consider the alternating direction method of multipliers to estimate the unknown parameters. Simulations conducted on two *ex vivo* lung datasets in which fluorescently labelled bacteria are present in the distal lung show good detection and correlation performance between bacteria counts identified by a trained clinician and those of the proposed method. A comparison with widely used spot-detection methods is also made, finding that the proposed algorithm showed superior detection performance.

Contributions

The main contributions of this work are fourfold and can be summarized as below.

1. A formulation of the problem of simultaneous sparse coding and outlier detection is proposed, and a numerical solver is provided.

2. The proposed algorithm can help to automatically detect labelled bacteria in optical endomicroscopy images. To the best of our knowledge, it is the first time this problem is addressed by a sparse coding approach.

3. We provide simulations using real datasets, in which we investigate different combinations of SmartProbes and bacterial concentrations including control cases in which no bacteria are present. Different SmartProbes that cause weak and strong fluorescence are tested.

4. We compare the results of the proposed model with bacteria annotations performed by a trained clinician, and three widely used spot detection algorithms, using both dot-annotation and count-annotation methods.

The remaining sections of this chapter are organized as follows. Section 5.2 reviews sparse representation modelling. Then, Section 5.3 provides a general approach for the dictionary learning problem, and the widely used dictionary learning methods. Section 5.4 formulates the problem of bacteria detection in OEM images using a sparse coding approach, followed by Sections 5.5 and 5.6, which summarize the proposed model and possible estimation strategy to estimate the unknown model parameters. Results and discussions on two real datasets of *ex vivo* ovine lungs instilled with bacteria are presented in Section 5.7. Intermediate conclusions are finally reported in Section 5.8.

5.2 Sparse Representation Modelling

In sparse representation, given a set of basis vectors, the goal is to decompose natural signals or images as linear combinations of only a few basis vectors. An important issue when formulating a problem as a sparse representation model is choosing the appropriate set of vectors over which each can be sparsely represented. In that sense, each vector is called an *atom*, and a collection of atoms is called a *dictionary*. In that sense, a signal, namely $\mathbf{y} \in \mathbb{R}^N$, can be described as $\mathbf{y} \approx \mathbf{D}\boldsymbol{\psi}$, where $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_M] \in \mathbb{R}^{N \times M}$ is a *dictionary* and $\boldsymbol{\psi} \in \mathbb{R}^M$ is sparse. The recovery of the sparse representation, which is called *sparse coding* can be obtained by solving the following minimization problem

$$\hat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}}{\text{minimize}} \quad \|\mathbf{D}\boldsymbol{\psi} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\psi}\|_0 \leq T_0, \tag{5.1}$$

where $\|\boldsymbol{\psi}\|_0$ is the ℓ_0 pseudo-norm that counts the number of non-zero entries of $\boldsymbol{\psi}$, and T_0 is the corresponding maximum number of non-zero entries. Problem (5.1) can be extended for a collection of signals stacked in columns as follows

$$\hat{\boldsymbol{\Psi}} = \underset{\boldsymbol{\Psi}}{\text{minimize}} \quad \|\mathbf{D}\boldsymbol{\Psi} - \mathbf{Y}\|_F^2 \quad \text{subject to} \quad \|\boldsymbol{\Psi}\|_{0,0} \leq LT_0, \tag{5.2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_L] \in \mathbb{R}^{N \times L}$ contains L signals stacked in columns, $\{\mathbf{Y}_i \in \mathbb{R}^N\}_{i=1}^L$, $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_L] \in \mathbb{R}^{M \times L}$ contains L sparse representations, $\{\boldsymbol{\Psi}_i \in \mathbb{R}^M\}_{i=1}^L$, $\|\boldsymbol{\Psi}\|_{0,0}$ counts the number of non-zero entries of $\boldsymbol{\Psi}$, and $\|\cdot\|_F$ is the Frobenius norm. This ℓ_0-norm minimization can be interpreted as finding the minimum quadratic fit obeying the sparsest solution constraint. This type of model is known as the single measurement vector (SMV) [57], as each signal is assumed to be a single measurement associated with a unique non-zero pattern of its sparse representation (i.e. unique combinations of atoms). Figure 5.1 illustrates an example of $\boldsymbol{\Psi}$ matrix in this case.

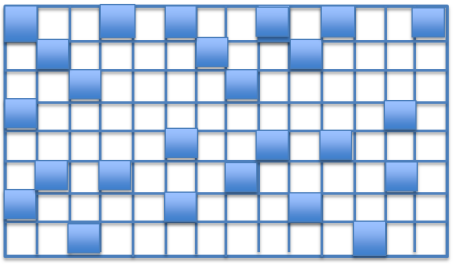


Figure 5.1 Schematic of sparse representation matrix; non-zeros, where dark squares represent non-zeros and white squares represent zeros.

Solving the optimization problem in Eq. (5.2) is an NP-hard problem due to the non-convexity and non-differentiability of the ℓ_0-norm. [57]. Approximate solutions are obtained using greedy algorithms such as matching pursuit (MP) or orthogonal matching pursuit (OMP) [57, 120]. A second class of methods relies on relaxation and replaces the ℓ_0-norm with an ℓ_1-norm convex programming problem, namely

$$\hat{\mathbf{\Psi}} = \underset{\mathbf{\Psi}}{\text{minimize}} \quad \|\mathbf{D\Psi} - \mathbf{Y}\|_F^2 \quad \text{subject to} \quad \|\mathbf{\Psi}\|_{1,1} \leq T_1, \tag{5.3}$$

where T_1 is the maximum sum of absolute entries of $\|\mathbf{\Psi}\|_{1,1}$. This makes the optimization problem a convex one that can be solved efficiently. Such methods are called basis pursuit (BP) [43].

5.3 Dictionary Learning

There are several ways to construct the dictionary, some of which are using pre-defined basis as discrete Fourier transform (DFT), discrete cosine transform (DCT), and/or discrete wavelets transform (DWT). However, pre-specified dictionaries do not necessarily describe well a given signal at hand. Hence, it is more appealing to learn the dictionary from a given set of training data, which has been found more efficient compared to using pre-defined ones [58,156]. Well known methods are K-singular value decomposition (K-SVD) [5], and the method of optimal directions (MOD) [61].

The joint estimation of $\mathbf{D}$ and $\mathbf{\Psi}$ in Eq. (5.2) can be written as unconstrained as follows

$$\{\hat{\mathbf{\Psi}}, \hat{\mathbf{D}}\} = \underset{\mathbf{\Psi}\in\mathcal{X}, \mathbf{D}\in\mathcal{D}}{\text{minimize}} \quad \|\mathbf{D\Psi} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{\Psi}\|_{p,p}, \tag{5.4}$$

where $\mathcal{D}$ and $\mathcal{X}$ are admissible sets of the dictionary and the sparse coefficient matrix respectively, λ is a positive regularization which weights the two terms of the objective functions, and p can be either $p = 0$ for ℓ_0 or $p = 1$ for ℓ_1, leading to joint sparseness prior distributions on $\mathbf{\Psi}$. The admissible set $\mathcal{D}$ is usually chosen as the set of dictionaries with unit column-norms. The aim is to find a dictionary $\mathbf{D}$, and a sparse representation matrix $\mathbf{\Psi}$ that achieves this minimization

problem. However, Eq.(5.4) is not jointly convex with respect to $(\mathbf{D}, \mathbf{\Psi})$, but with respect to each of the variables $\mathbf{D}$ and $\mathbf{\Psi}$ when the other one is fixed. Hence, a general approach to solve Eq. (5.4) would be to use alternating minimization, as shown in Algorithm 9.

Algorithm 9 A General Approach for Dictionary Learning

1: **Fixed input parameters**: Data matrix $\mathbf{Y}$, total number of iterations t, Regularization parameter λ.
2: **Initializations:** Set $k = 0$, $\mathbf{\Psi}^{(0)}$ and $\mathbf{D}^{(0)}$
3: **Repeat** $(1 \leq k \leq t)$
- $\mathbf{\Psi}^{(k+1)} \leftarrow \underset{\mathbf{\Psi} \in \mathcal{X}}{\text{minimize}} \left\|\mathbf{D}^{(k)}\mathbf{\Psi} - \mathbf{Y}\right\|_F^2 + \lambda \|\mathbf{\Psi}\|_{p,p}$ (3a),
- $\mathbf{D}^{(k+1)} \leftarrow \underset{\mathbf{D} \in \mathcal{D}}{\text{minimize}} \left\|\mathbf{D}\mathbf{\Psi}^{(k+1)} - \mathbf{Y}\right\|_F^2$ (3b),
- $k \leftarrow k + 1$.

4: **until** some stopping criterion is satisfied.

Equation (3a) in Algorithm 9 is indeed a sparse coding problem for which many algorithms have been proposed as previously mentioned in Section 5.2. The main difference between various dictionary learning algorithms is their approach to updating the dictionary (Eq.(3b) in Algorithm 9). Two well-known algorithms are K-SVD [5], and MOD [61]. MOD finds the unconstrained minimum of the dictionary update problem, resulting in a closed-form solution, which is simply a least squares problem, and then projects it onto $\mathcal{D}$ by normalizing its columns. In this way, MOD updates all atoms at once. K-SVD, on the other hand, uses a sequential approach which is similar to the K-means clustering algorithm in order to update the atoms one by one.

The next section provides a formulation of the bacteria detection problem using sparse representation, when the dictionary $\mathbf{D}$ is known (estimated a priori using one of the methods mentioned above).

5.4 Sparse Representation Formulation to the Bacteria Detection Problem

Fig. 5.2 shows a typical sequence of OEM frames with labelled bacteria annotated by a trained clinician. We can observe that labelled bacteria appear as bright dots overlaid across a network of very bright elastin strands in these images. In order to identify and quantify these bright spots, we consider the following formulation.

Given a test image $\mathcal{I} \in \mathbb{R}^{m \times n}$, a data matrix $\mathbf{Y} \in \mathbb{R}^{P \times L}$ is formed by splitting the image into a set of overlapping patches each of size $N_p \times N_p$, namely $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_L]$. The data matrix $\mathbf{Y}$ can then be well approximated by a sparse linear model, excluding a small number of pixels - the outliers - which significantly deviate from this model. The collection $\mathbf{Y}$ is described as follows

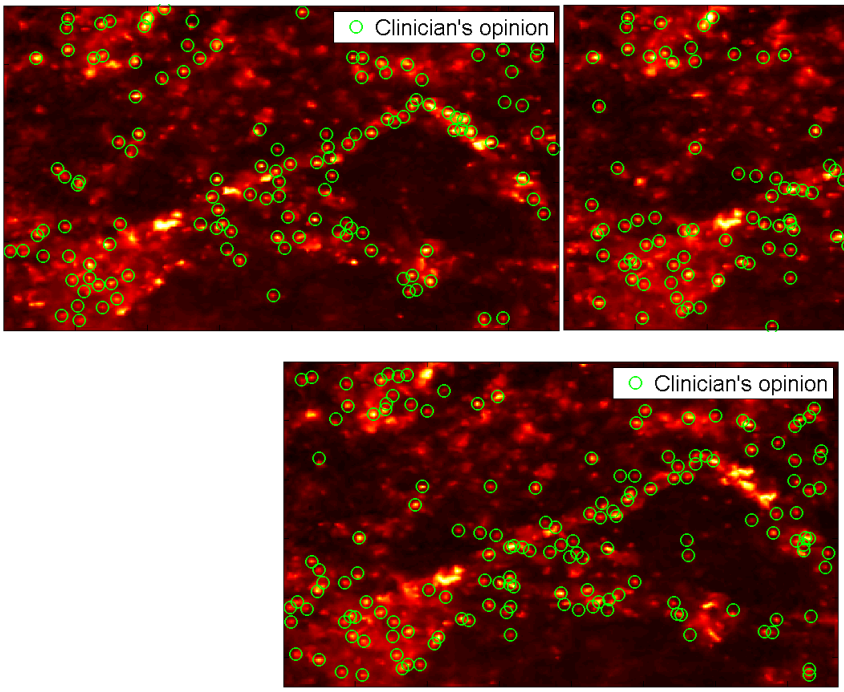


Figure 5.2 A sequence of OEM frames with bacteria shown within circles annotated by a trained clinician.

$$\mathbf{Y} = \mathbf{D}\boldsymbol{\Psi} + \mathbf{R} + \mathbf{W}, \tag{5.5}$$

where $\mathbf{D} \in \mathbb{R}^{P\times K}$ is the dictionary and is assumed known, $\boldsymbol{\Psi} \in \mathbb{R}^{K\times L}$ is the sparse coefficient matrix, $\mathbf{R} \in \mathbb{R}^{P\times L}$ has few non-zero elements that equal the deviation of each outlier from the sparse representations model, and $\mathbf{W} \in \mathbb{R}^{P\times K}$ is a low-energy noise component, which is assumed to be Gaussian.

As the number of fibre cores in each patch in the image might be different, in this work, we consider processing the full interpolated OEM image rather than just the central fibre core intestines. This will make the formation of the data matrix $\mathbf{Y}$ easier. Moreover, considering the full interpolated image will make the characterization of background image structures more robust than representing them with spatially irregular samples.

The objective here is to estimate the outlier matrix $\mathbf{R}$ in Eq.(5.5), but the sparse coefficients matrix $\boldsymbol{\Psi}$ is also unknown. Thus we estimate jointly $\mathbf{R}$ and $\boldsymbol{\Psi}$ from the observation matrix $\mathbf{Y}$. To solve this problem, we propose an optimization-based method to estimate the unknown parameters.

5.5 Proposed Model

This section introduces an optimization-based method to estimate the unknown model parameter $\mathbf{R}$ and $\boldsymbol{\Psi}$ in Eq.(5.5), which can be estimated by solving the following constrained minimization

problem

$$\begin{aligned} \{\hat{\mathbf{\Psi}}, \hat{\mathbf{R}}\} = \underset{\mathbf{\Psi},\mathbf{R}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{\Psi} - \mathbf{R}\|_F^2, \\ \text{subject to} \quad & \|\mathbf{\Psi}\|_{1,1} \leq T_1, \\ & \|\mathbf{R}\|_{1,1} \leq S_1, \end{aligned} \tag{5.6}$$

where $\|\mathbf{\Psi}\|_{1,1} = \sum_j \|\mathbf{\Psi}(j,:)\|_1$, similarly $\|\mathbf{R}\|_{1,1} = \sum_j \|\mathbf{R}(j,:)\|_1$, and S_1 is the maximum sum of absolute entries of $\mathbf{R}$ (i.e. outliers). Problem (5.6) encourages a solution in which $\mathbf{\Psi}$ is sparse. However, for the outliers that cannot be represented exclusively by $\mathbf{D}$, it permits non-zero entries in $\mathbf{R}$. The constraints ensure at most T_0 and S_0 non-zero entries in $\mathbf{\Psi}$ and $\mathbf{R}$ respectively. The constrained version of Eq.(5.6) can be written in an unconstrained form as follows

$$\{\hat{\mathbf{\Psi}}, \hat{\mathbf{R}}\} = \underset{\mathbf{\Psi},\mathbf{R}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{\Psi} - \mathbf{R}\|_F^2 + \alpha\|\mathbf{\Psi}\|_{1,1} + \beta\|\mathbf{R}\|_{1,1}, \tag{5.7}$$

where α and β are two positive scalars controlling the degree of sparsity of $\mathbf{\Psi}$ and $\mathbf{R}$ respectively. The next section proposes an estimation strategy to solve Eq. (5.7).

5.6 Estimation Strategy

The optimization problem in Eq. (5.7), although convex, is very hard to solve owing to the non-smooth terms. The core idea is to convert the unconstrained minimization problem in Eq. (5.7) into another constrained one via variable splitting by the application of a variable splitting operation. Finally, the obtained constrained problem is dealt with using the alternating direction method of multipliers (ADMM) [3], which belongs to the family of augmented Lagrangian techniques [138]. By a careful choice of the new variables, the initial problem is converted into a sequence of much simpler problems, which can be solved iteratively. The choice of ADMM here is due to the following reasons: (i) it is suitable for our problem format, (ii) it has proven convergence properties, and (iii) it leads to a simple coordinate-descent structure.

In this problem, we introduce a new variable $\mathbf{Z}$ for the regularization function in $\mathbf{\Psi}$ in order to decouple it from the data fidelity term. Therefore, the constrained version of problem (5.7) can be written as follows

$$\begin{aligned} \{\hat{\mathbf{\Psi}}, \hat{\mathbf{R}}\} = \underset{\mathbf{\Psi},\mathbf{R}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{\Psi} - \mathbf{R}\|_F^2 + \alpha\|\mathbf{Z}\|_{1,1} + \beta\|\mathbf{R}\|_{1,1}, \\ \text{subject to} \quad & \mathbf{Z} = \mathbf{\Psi}. \end{aligned} \tag{5.8}$$

The augmented Lagrangian corresponding to Problem (5.8) can be written as follows

$$\mathcal{L}(\mathbf{\Psi}, \mathbf{R}, \mathbf{Z}, \mathbf{M}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{\Psi} - \mathbf{R}\|_F^2 + \alpha\|\mathbf{Z}\|_{1,1} + \beta\|\mathbf{R}\|_{1,1} + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{\Psi} - \mathbf{M}\|_F^2, \tag{5.9}$$

where $\mathbf{M}$ is the Lagrange multiplier corresponding to the splitting, and $\mu > 0$ is a constant. The ADMM algorithm corresponding to problem Eq.(5.8) (also Eq.(5.7)) is shown in Algorithm 10. This algorithm considers an iterative way to optimize minimization problems of each parameter when the others are fixed, then updates the Lagrange multipliers. The rationale here is that each step of this alternating minimization may be much easier than solving the original unconstrained problem Eq.(5.7). During each step of this iterative scheme, $\mathcal{L}$ is optimized w.r.t. $\mathbf{\Psi}$ (step 3), $\mathbf{R}$ (step 4) and $\mathbf{Z}$ (step 5), and then the Lagrange multiplier is updated (step 6). The algorithm terminates when a stopping criterion is satisfied, as will be mentioned below.

Algorithm 10 Patch-Based Bacteria Detection Using ADMM - Version I

1: set $k = 0$, choose $\mu > 0, \mathbf{\Psi}^{(0)}, \mathbf{R}^{(0)}, \mathbf{Z}^{(0)},$ and $\mathbf{M}^{(0)}$
2: **repeat** $(k \leftarrow k + 1)$
3: $\mathbf{\Psi}^{(k+1)} \leftarrow \text{minimize}_{\mathbf{\Psi}}\, \mathcal{L}(\mathbf{\Psi}, \mathbf{R}^k, \mathbf{Z}^k, \mathbf{M}^k)$
4: $\mathbf{R}^{(k+1)} \leftarrow \text{minimize}_{\mathbf{R}}\, \mathcal{L}(\mathbf{\Psi}^{k+1}, \mathbf{R}, \mathbf{Z}^k, \mathbf{M}^k)$
5: $\mathbf{Z}^{(k+1)} \leftarrow \text{minimize}_{\mathbf{Z}}\, \mathcal{L}(\mathbf{\Psi}^{k+1}, \mathbf{R}^{k+1}, \mathbf{Z}, \mathbf{M}^k)$
6: **Update Lagrange multipliers:** $\mathbf{M}^{(k+1)} \leftarrow \mathbf{M}^{(k)} - \left(\mathbf{Z}^{(k+1)} - \mathbf{\Psi}^{(k+1)}\right)$
7: **until** some stopping criterion is satisfied.

We now detail each step of Algorithm 10 as follows.

Step 3: Updating the sparse codes $\mathbf{\Psi}$

The goal of this step is to determine the value of the variable $\mathbf{\Psi}$ at each iteration. Given that we run an optimization over the variable $\mathbf{\Psi}$, the terms of the objective function in Eq. (5.8) which do no contain this variable are not taken into account. The reduced optimization function becomes, then,

$$\mathbf{\Psi}^{(k+1)} \leftarrow \underset{\mathbf{\Psi}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{\Psi} - \mathbf{R}^{(k)}\|_F^2 + \frac{\mu}{2}\|\mathbf{Z}^{(k)} - \mathbf{\Psi} - \mathbf{M}^{(k)}\|_F^2, \tag{5.10}$$

the solution to which is simply

$$\mathbf{\Psi}^{(k+1)} \leftarrow \left(\mathbf{D}^T\mathbf{D} + \mu^{(k)}\mathbf{I}\right)^{-1}\left(\mathbf{D}^T(\mathbf{Y} - \mathbf{R}^{(k)}) + \mu^{(k)}(\mathbf{Z}^{(k)} + \mathbf{M}^{(k)})\right). \tag{5.11}$$

Step 4: Updating the outlier matrix R

On the other hand, step 4 of Algorithm 10 computes the value of the variable $\mathbf{R}$ at the current iteration. The optimization problem to be solved is

$$\mathbf{R}^{(k+1)} \leftarrow \underset{\mathbf{R}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{\Psi}^{(k+1)} - \mathbf{R}\|_F^2 + \beta\|\mathbf{R}\|_{1,1}, \tag{5.12}$$

the solution to which is the well-known *soft* threshold [49]

$$\mathbf{R}^{(k+1)} \leftarrow \text{soft}\left(\mathbf{Y} - \mathbf{D}\boldsymbol{\Psi}^{(k+1)}, \beta\right), \tag{5.13}$$

where $\text{soft}(e, \tau)$ denotes the component-wise application of the soft-threshold function $\text{sign}(e)\max\{|e| - \tau, 0\}$.

Step 5: Updating the auxiliary matrix Z

Similar to updating the outlier matrix $\mathbf{R}$, updating $\mathbf{Z}$ is reduced to solving the following minimization problem

$$\mathbf{Z}^{(k+1)} \leftarrow \underset{\mathbf{Z}}{\text{minimize}} \quad \frac{\mu}{2}\|\mathbf{Z} - \boldsymbol{\Psi} - \mathbf{M}\|_F^2 + \alpha\|\mathbf{Z}\|_{1,1}, \tag{5.14}$$

the solution to which is also the *soft* threshold function

$$\mathbf{Z}^{(k+1)} \leftarrow \text{soft}\left(\boldsymbol{\Psi}^{(k+1)} - \mathbf{M}^{(k)}, \frac{\alpha}{\mu}\right). \tag{5.15}$$

The final version of Algorithm 10 is shown in Algorithm 11. The parameter $\mu > 0$ is updated within the algorithm to keep the primal and dual residual norms within a factor of $\rho = 10$ of one another. This has been proven to provide fast convergence [32]. The stopping criterion we use is $\left(\|\mathbf{Z}^{(k)} - \boldsymbol{\Psi}^{(k)}\|_F + \mu\|\mathbf{M}^{(k)} - \mathbf{M}^{(k+\rho)}\|_F\right) \leq \epsilon$, which is the sum of the primal and dual residuals, where $\epsilon = \sqrt{P \times L} \times 10^{-6}$ [3].

Algorithm 11 Patch-Based Bacteria Detection Using ADMM - Version II

1: set $k = 0$, choose $\mu > 0, \boldsymbol{\Psi}^{(0)}, \mathbf{R}^{(0)}, \mathbf{Z}^{(0)},$ and $\mathbf{M}^{(0)}$
2: **repeat** $(k \leftarrow k + 1)$
3: $\boldsymbol{\Psi}^{(k+1)} \leftarrow \left(\mathbf{D}^T\mathbf{D} + \mu^{(k)}\mathbf{I}\right)^{-1}\left(\mathbf{D}^T(\mathbf{Y} - \mathbf{R}^{(k)}) + \mu^{(k)}(\mathbf{Z}^{(k)} + \mathbf{M}^{(k)})\right)$,
4: $\mathbf{R}^{(k+1)} \leftarrow \text{soft}\left(\mathbf{Y} - \mathbf{D}\boldsymbol{\Psi}^{(k+1)}, \beta\right)$,
5: $\mathbf{Z}^{(k+1)} \leftarrow \text{soft}\left(\boldsymbol{\Psi}^{(k+1)} - \mathbf{M}^{(k)}, \frac{\alpha}{\mu}\right)$,
6: $\mathbf{M}^{(k+1)} \leftarrow \mathbf{M}^{(k)} - \left(\mathbf{Z}^{(k+1)} - \boldsymbol{\Psi}^{(k+1)}\right)$
7: $\mu^{(k+1)} \leftarrow \rho\mu^{(k)}$
8: **until** some stopping criterion is satisfied.

5.7 Simulations Using Real Datasets

5.7.1 Datasets

The proposed algorithm was assessed using two datasets of *ex vivo* ventilated whole ovine lungs with bacteria present. Dataset I contains seven videos assessing a combination of fluorescent dyes and bacterial types, including control segments. It contains (i) three videos of ovine lungs instilled

with *Methicillin-sensitive Staphylococcus aureus* (*MSSA*) stained with a commercially available laboratory dye (PKH67, Sigma-Aldrich), a highly fluorescent cell membrane dye, (ii) two videos of ovine lungs instilled with bacteria (gram-positive *MSSA* and gram-negative *Pseudomonas* PA3284) stained in situ with an in-house bacterial detection SmartProbe [6], and (iii) two videos of ovine lungs without the presence of any bacteria. Videos 1 to 5 were instilled with a single concentration of bacteria, equivalent to Optical Density (OD595nm) of 2.

Video	**Number of frames**	**Bacteria concentration (OD)**	**Fluorophore**	**Bacteria**
1	26	2	PKH	*Staphylococcus aureus*
2	19			
3	13			
4	32		SmartProbe	*Pseudomonas aeruginosa*
5	19			*Staphylococcus aureus*
6	12	NA	NA	Control
7	12			

Table 5.1 Description of dataset I.

Dataset II contains four videos, each with an increasing bacterial concentration (OD595nm 0.004, 0.04, 0.4, 4), all labelled with an in-house bacterial detection SmartProbe. This dataset is considered to compare the annotations performed by the clinician and the results of the algorithm. As it will ensure that as the concentration increases, the counts of the clinician and of the algorithm also increase. Tables 5.1 and 5.2 summarise the details of Datasets I and II respectively.

Video	**# of frames**	**Bacteria concentration (OD)**	**Fluorophore**	**Bacteria**
1	14	0.004	SmartProbe	*Pseudomonas aureus*
2	14	0.04		
3	15	0.4		
4	15	4		

Table 5.2 Description of dataset II.

The Cellvizio fibred confocal OEM imaging platform, along with a 1.4 mm, 12,105 core fibre bundle (Mauna Kea Technologies, Paris, France) was used to acquire all data in this study. Image sequences of size 274×384 pixels ($306\,\mu m \times 429\,\mu m$) were captured at 12 frames per second.

Random frames that are representative of each of the entire video sequences are chosen from each of the eight videos of Dataset I and the four videos of Dataset II by a trained clinician. These comprise 133 image frames for Dataset I, and 58 frames for Dataset II as described in Tables 5.1 and 5.2 respectively. In each frame, a trained clinician marked the co-ordinates of phenomena that

are thought with high confidence to be bacteria. Ambiguous points are ignored. Due to the size of the imaged bacteria (<3 µm), a bright fluorescent 'dot' is identified as a bacterium if one or two cores present higher intensities than the surrounding area.

In order to unify the dynamic range of the frames, every one is normalized to the $0 : 255$ intensity range.

5.7.2 Dictionary Learning for Bacterial Detection

In this spot detection problem, each dataset is split into training, testing phases. In the training phase, one dictionary is learned for each dataset from its corresponding videos; namely $\mathbf{D}_1$ for Dataset I, and $\mathbf{D}_2$ for Dataset II. Every set of frames in each video has a certain elastin and collagen pattern. Hence, one frame from each set is chosen as a representative. This yielded 12 frames from Dataset I and 17 frames from Dataset II (See Figures 5.3 and 5.4). Features are then extracted from each frame by dividing it into square overlapping patches of fixed size. In this work, we employed a 27×27 window size with 50% overlap. The patches that are annotated by the clinician as having bacteria are then excluded from the training dataset. Each of the remaining features is then vectorized into one column resulting in a training data matrix $\mathbf{Y} \in \mathbb{R}^{729\times 3553}$ for Dataset I and $\mathbf{Y} \in \mathbb{R}^{729\times 7087}$ Dataset II. The MOD dictionary learning method is then applied to train the dictionaries. Different atom numbers are tested including 100, 200 and 300 atoms. The K-SVD algorithm is also tested but provided similar results to the MOD, thus the results are not provided here. Figures 5.5 and 5.6 show the 100 dictionary atoms from each of Datasets I and II.

5.7.3 Algorithm evaluation

In the testing phase, after the dictionaries have been learned, Algorithm 11 is run for each of the remaining frames which are 111 for Dataset I and 41 for Dataset II, yielding the estimated outlier matrix $\hat{\mathbf{R}}$. The final outlier image is then reconstructed using these overlapping patches by averaging their intensities. The outlier image is then normalized to $[0, 1]$ range and thresholded (by ℓ_d), while pixels that exceed this threshold value are counted as a potential bacteria. Since each bacterium anticipates a set of pixels, each group of connected detections is grouped and counted as a single detection. Once these connected detections are identified, they are then replaced by a single detection at the mean of their locations, which gives the final number of detected bacteria in each frame.

Due to the absence of true negative (TN) counts in our problem, we consider precision-recall curves in order to assess the bacteria detection performance, in which the reference is the clinician's annotations. Precision-recall curves are plots of precision versus recall at different cut-off thresholds (different ℓ_d) for the resulting outlier amplitude image. The precision (also equivalent to positive predictive value) and the recall (also equivalent to sensitivity or true positive rate) can be calculated

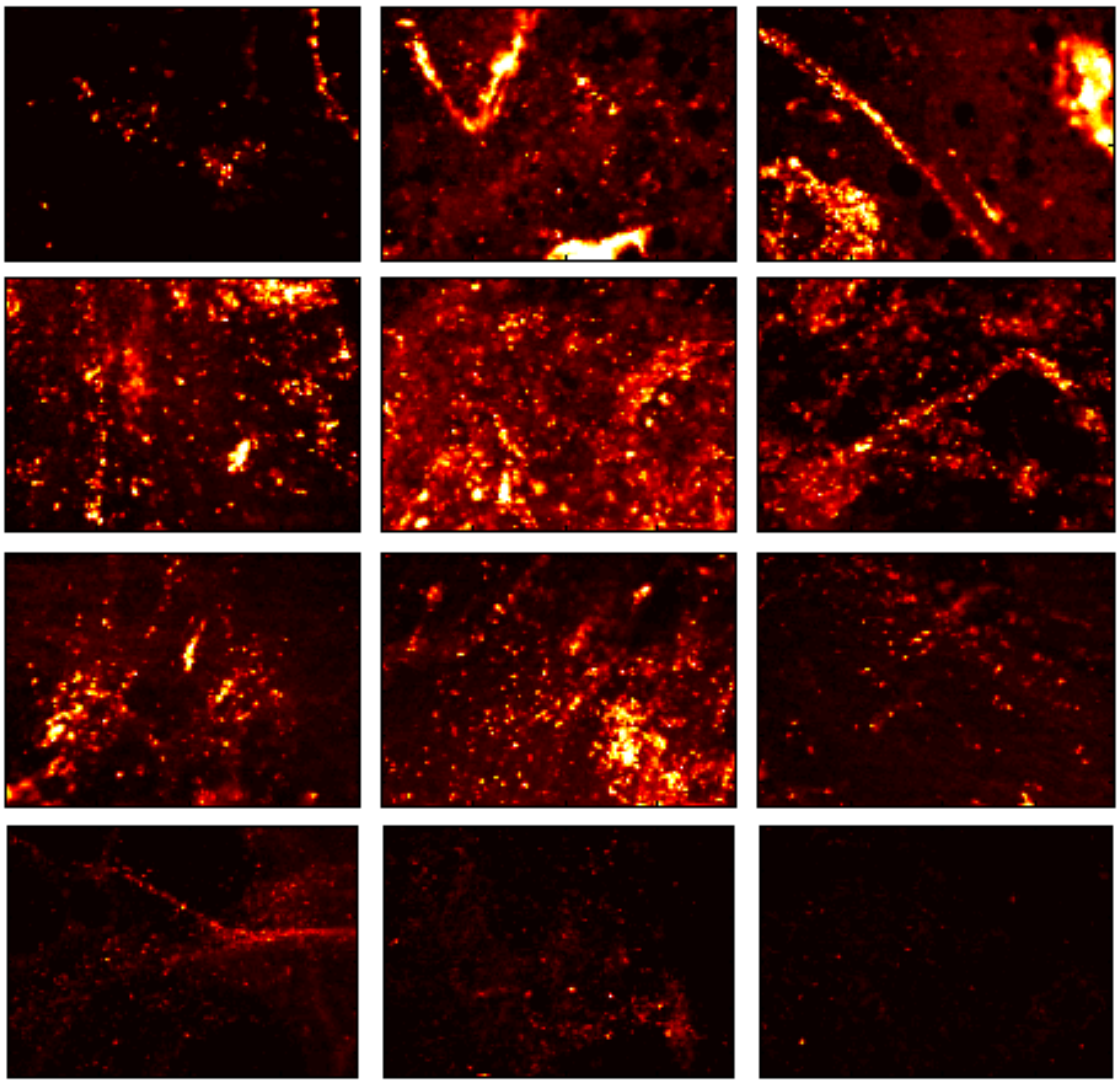


Figure 5.3 Selection of frames selected to train the dictionary for Dataset I. Note that the patches that have bacteria are excluded from the training process.

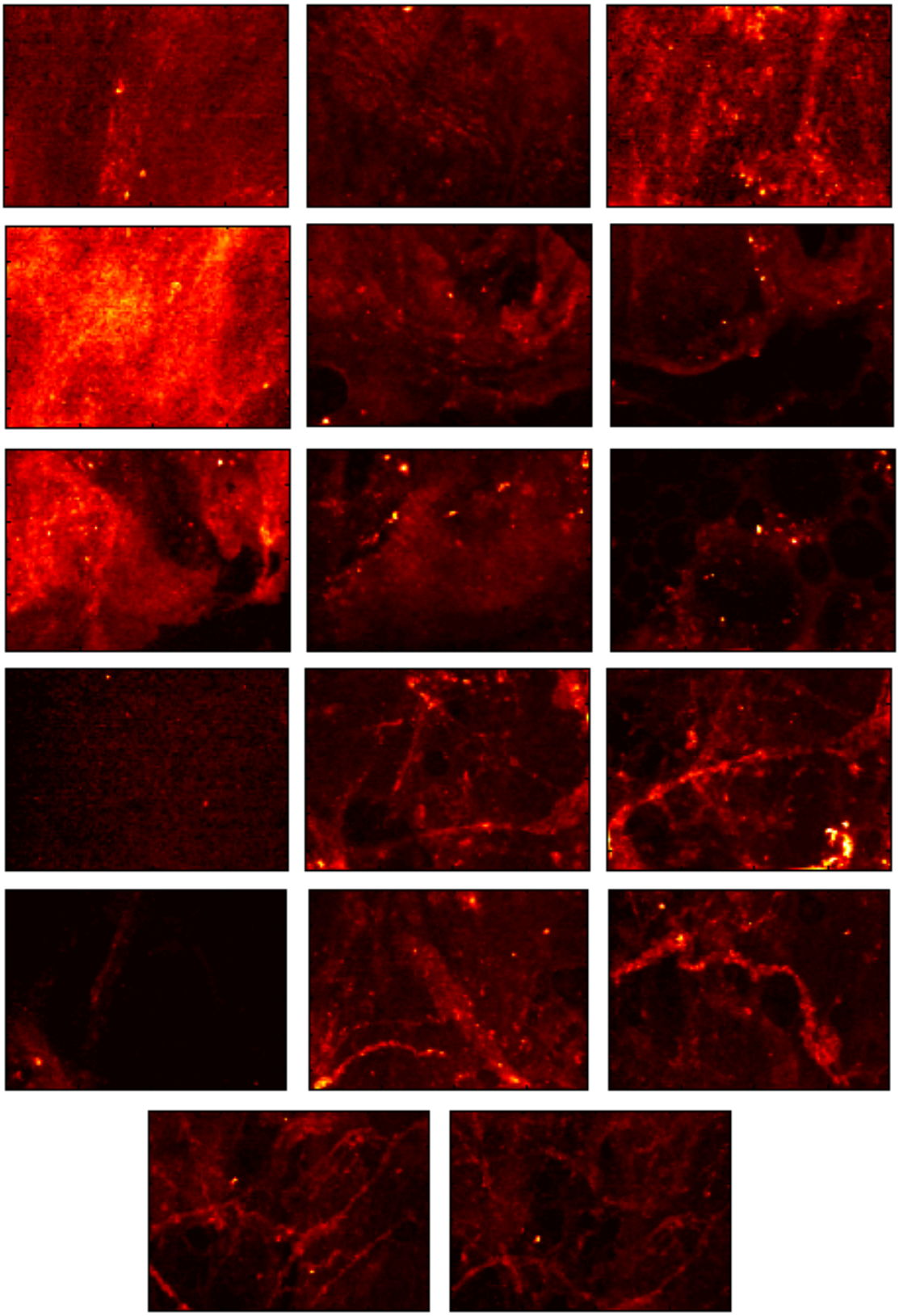


Figure 5.4 Selection of frames selected to train the dictionary for Dataset II. Note that the patches that have bacteria are excluded from the training process.

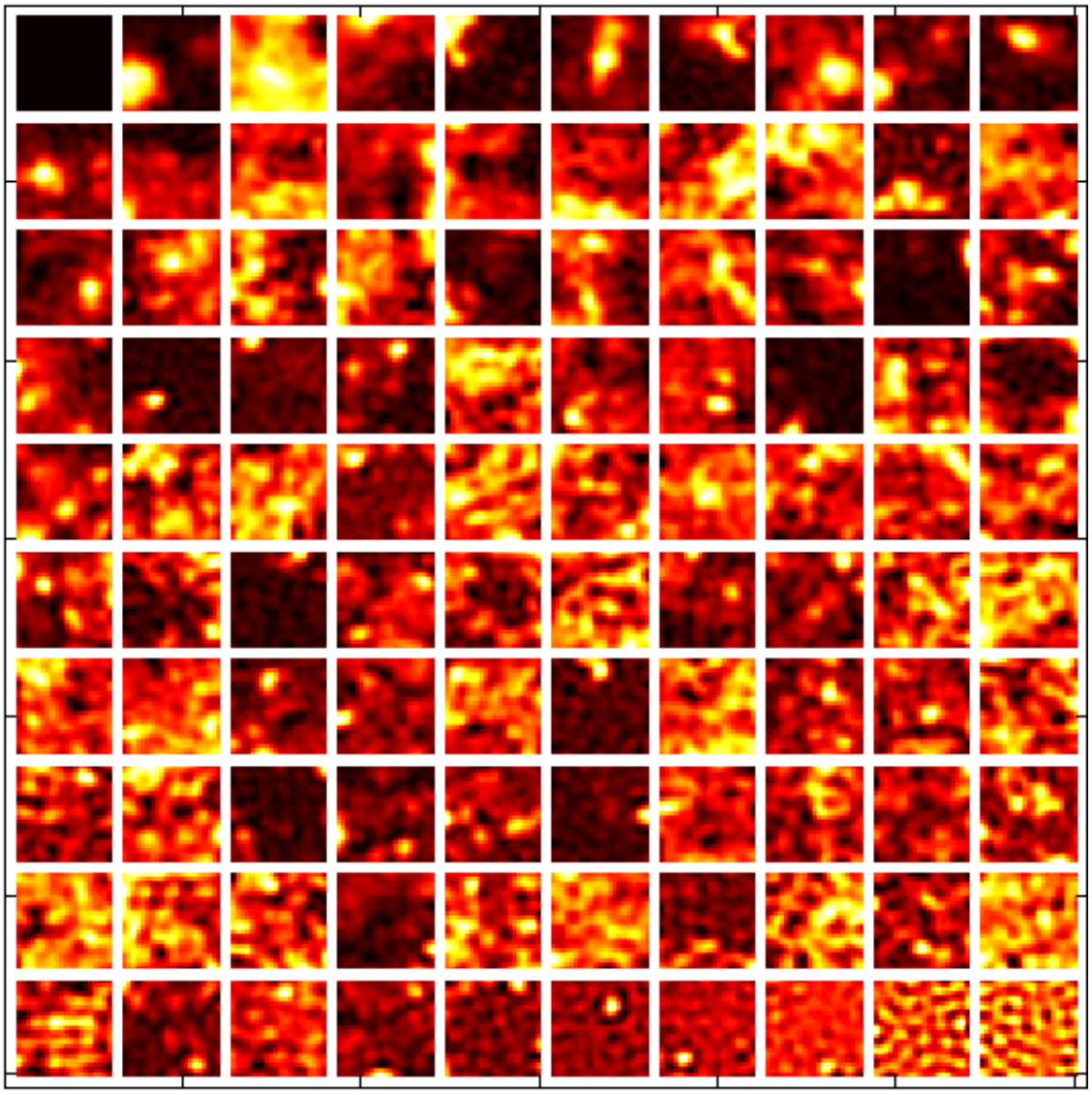


Figure 5.5 Dictionary atoms learned from Dataset I.

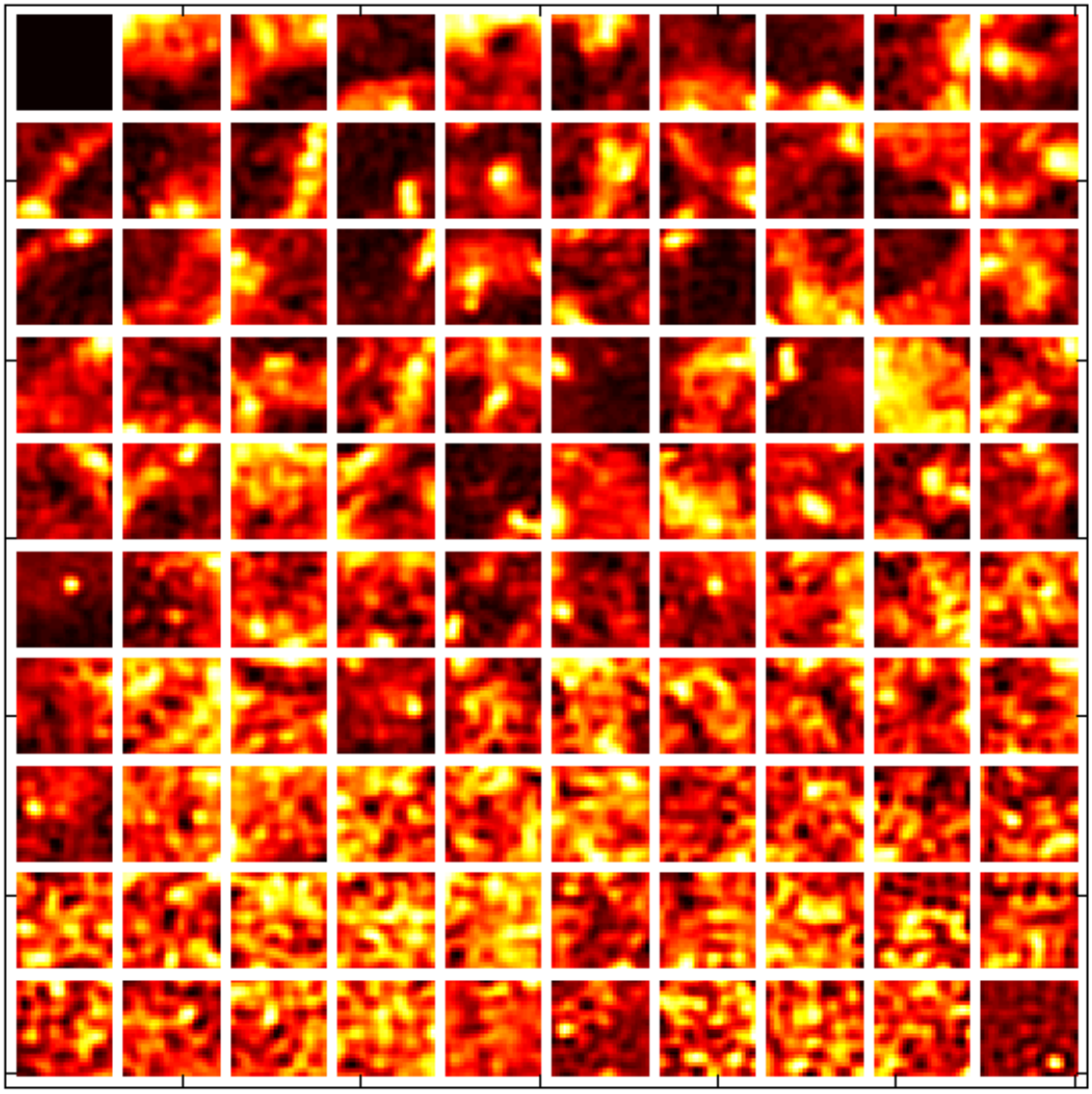


Figure 5.6 Dictionary atoms learned from Dataset II.

as follows

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \qquad \text{Recall} = \frac{\text{FP}}{\text{FP} + \text{FN}},$$

where TP, FN, and FP refer to the number of true positives, false negatives, and false positives respectively. Given the pixel locations where a bacterium has been detected and annotated by the clinician, we defined a disk of radius r pixels around each clinician's annotations [121], and we consider

- Any detection that exists within the disk to be a match (TP).
- Any detection outside any of the disks to be FP.
- Any clinician's annotation that does not match with any of the algorithm detection to be FN.

The mean distance in pixels between fibre cores is approximately 4.1 pixels. A radius of $r = 10$ pixels is considered.

We test different parameters for evaluating the performance of the proposed algorithm. First, we fix the regularization parameter α corresponding to the sparse representation matrix $\mathbf{\Psi}$ to $\alpha = 1 \times 10^{-5}$, and test different outlier regularization parameter values (β). Second, we test different numbers of atoms (K) within the learned dictionary. Finally we vary the outlier amplitude image threshold (ℓ_d) between 0 and 1, and construct the precision-recall curves accordingly. Statistical comparison of bacterial counts (count-annotation) and detections (dot-annotation) performed by the trained clinicians and the algorithm output is then considered after choosing the best combination of the parameters described above.

5.7.4 Results and Discussion

Dot-annotation effect

Figure 5.7 presents precision-recall curves for different outlier regularization parameters (β), different numbers of dictionary atoms (K) and different smartprobes (represented by video range) in Dataset I. It can be observed that a broad range of outlier regularization parameters provides very similar precision-recall curves. Figure 5.8, shows a plot of different smartprobes (represented by video ranges) versus different numbers of atoms and the maximum achieved area under curve (AUC) reported in Figure 5.7. The maximum achieved AUC is 0.754, 0.8 and 0.22 for videos 1:3, 4:5 and 6:7 respectively, and the average AUC of the whole dataset is 0.61. It can be noted that the bacteria detection performance is enhanced when increasing the number of dictionary atoms. Although a strong smartprobe which produces high fluorescence signals is used for videos 1:3, the

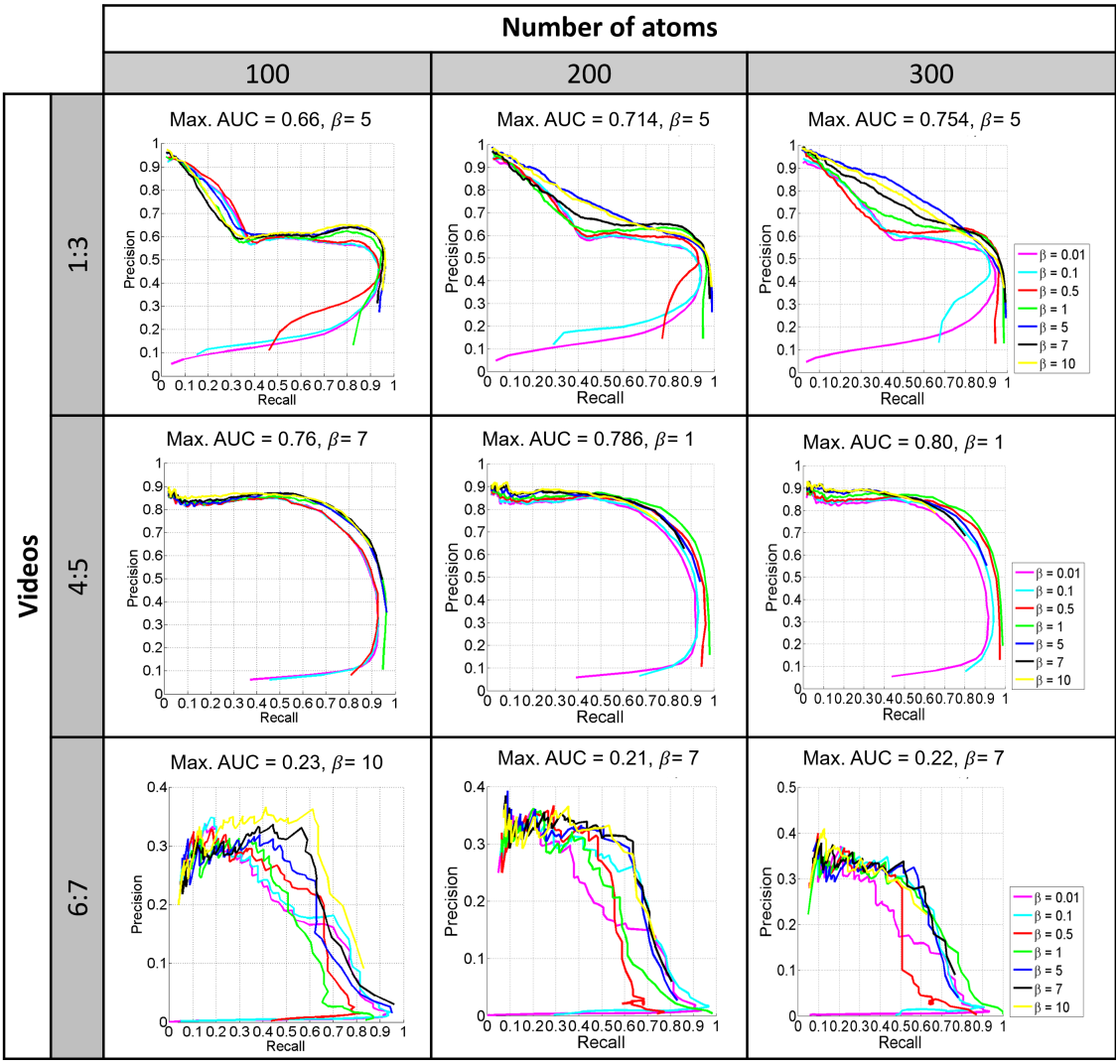


Figure 5.7 Precision-recall curves for different number of dictionary atoms and different values of outlier regularization parameter β for Dataset I.

reported AUC is slightly lower than that for videos 3:4, for which a weaker smartprobe is used. This is because videos 4 and 5 have less elastin and collagen structures, and hence there is a low chance of getting false positive detections. Regarding the control cases (videos 6 and 7), it can be observed that the optimal regularization parameter β is always higher than that of the bacteria stained videos (videos 1:3 and 4:5), which in turn promotes more outlier sparseness and hence fewer counts. Moreover, the AUCs of these videos are lower than those of videos 1:3 and 4:5, as they are not stained by fluorophores and hence makes the fluorescence of bacteria weaker and more difficult to discriminate, stressing on the need on SmartProbes for bacterial detection.

Figure 5.9, on the other hand, presents the resulting precision-recall curves achieved per video (or concentration) in Dataset II when different outlier regularization parameter values (β) and different numbers of dictionary atoms are tested. There is a general trend that when the bacteria

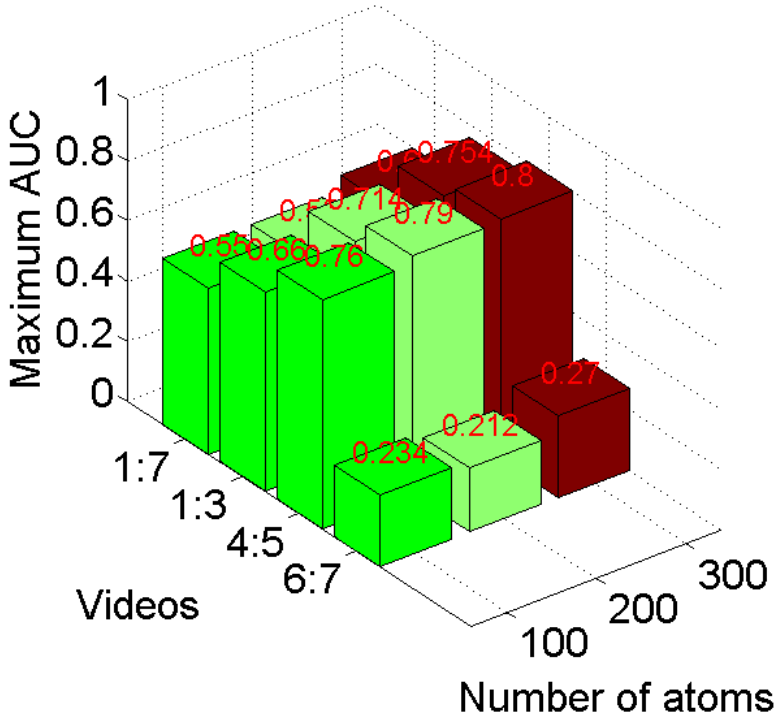


Figure 5.8 Plot of maximum achieved AUC reported from Figure 5.7 for Dataset I. The maximum values are provided above each bar.

concentration increases, the outlier regularization parameter β value slightly decreases. Figure 5.10 shows a plot of the three concentrations (represented by video numbers) versus different numbers of atoms and the maximum achieved AUC in Figure 5.9. We can note that there is not much difference in the achieved maximum AUC for the two tested dictionary atom numbers.

Count-annotation effect

For Dataset I, the algorithm counts are compared with the clinician counts in each frame as shown in Fig. 5.11, at a cut-off threshold of 0.07 to the outlier amplitude images, which corresponds to precision of 50% and recall of 86.12%. We can observe an almost linear relationship between the clinician counts and algorithm counts. Moreover the empirical correlation between the manually and automatically detected anomalies is 0.823. Moreover, for videos 1, 2 and 3 in which a highly fluorescent SmartProbe is used, and videos 4 and 5 in which an in-house SmartProbe which produces weaker fluorescence signals is used, a similar trend is observed between the clinician and the algorithm. This also applies to the type of bacteria the samples are labelled with. Videos 6 and 7, which are controls, demonstrate minimal counts using both the clinician and the algorithm, which reflects the ability of the algorithm to differentiate bacterial loads from control.

Similarly, for Dataset II, we tested the clinician-algorithm counts for different cut-off thresholds ℓ_d, ranging between $\ell_d = 0.05$ and $\ell_d = 0.08$, which corresponds to total sensitivity of 82.03% :

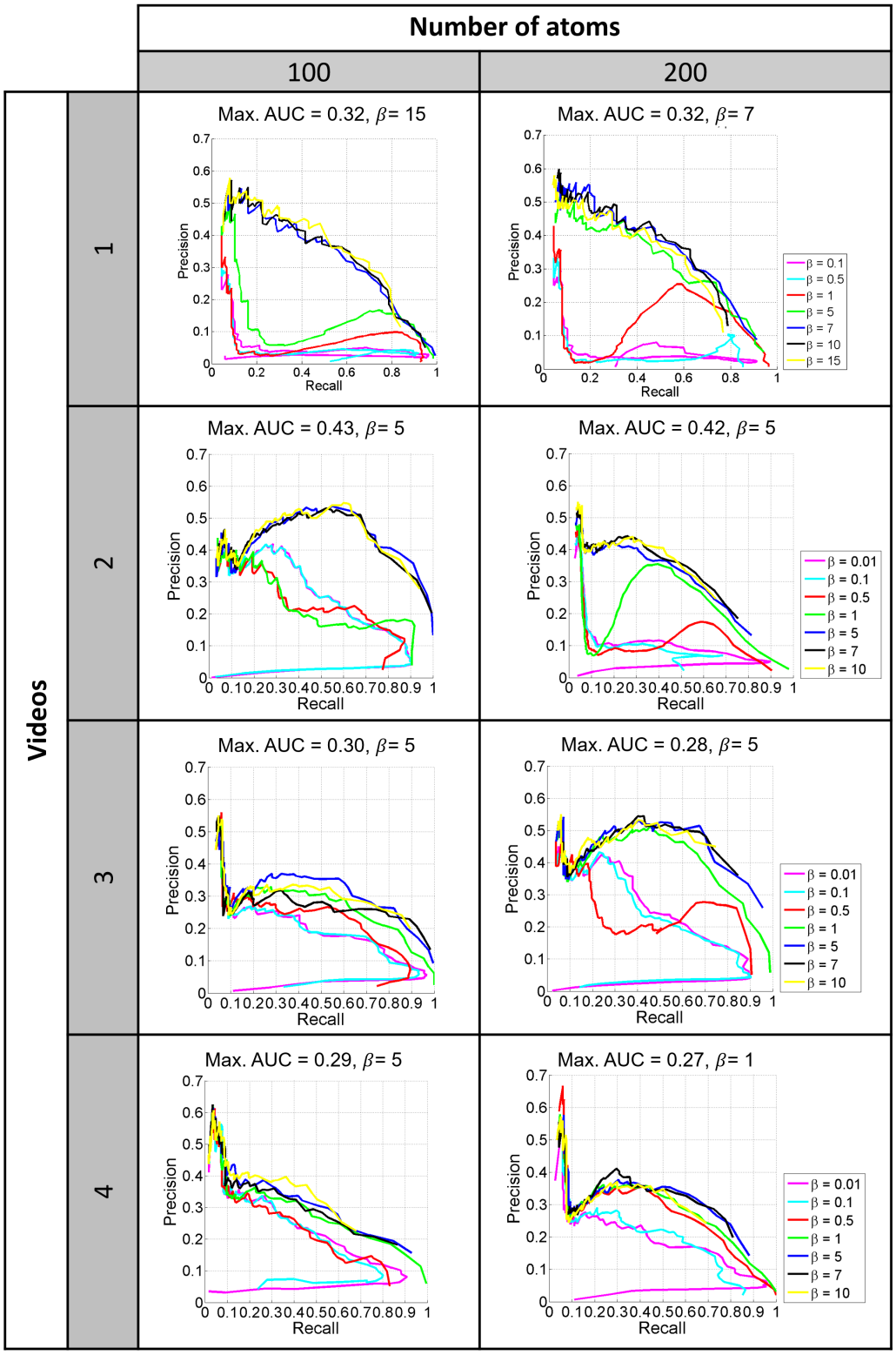


Figure 5.9 Precision-recall curves for different numbers of dictionary atoms and different values of outlier regularization parameter β for Dataset II.

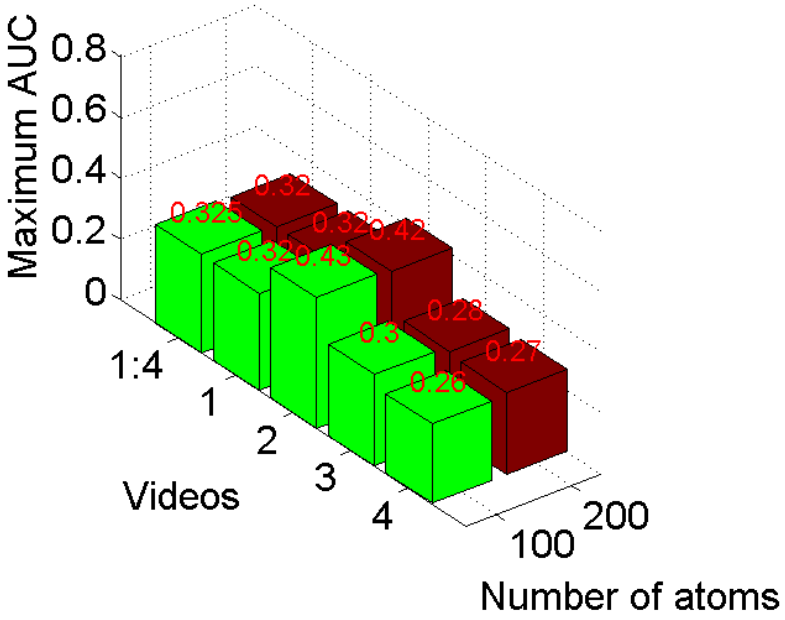


Figure 5.10 Plot of maximum achieved AUC reported from Figure 5.9 for Dataset II. The maximum values are provided above each bar.

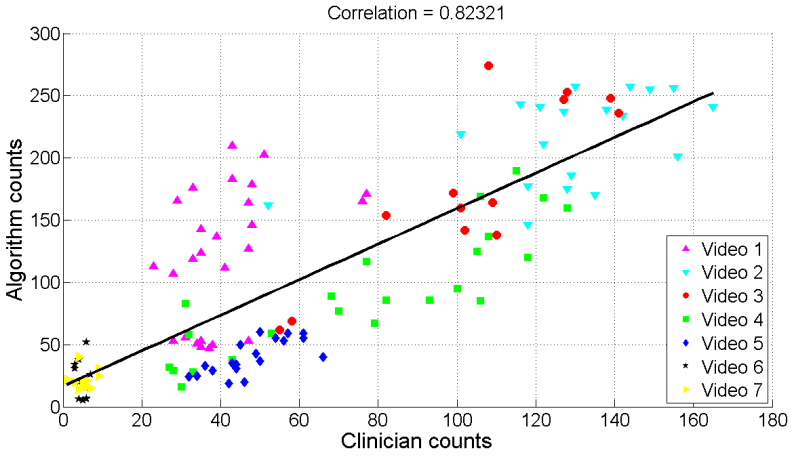


Figure 5.11 Plot of clinician bacteria count versus algorithm bacteria count for dataset I. Dots correspond to frames, and colours correspond to videos.

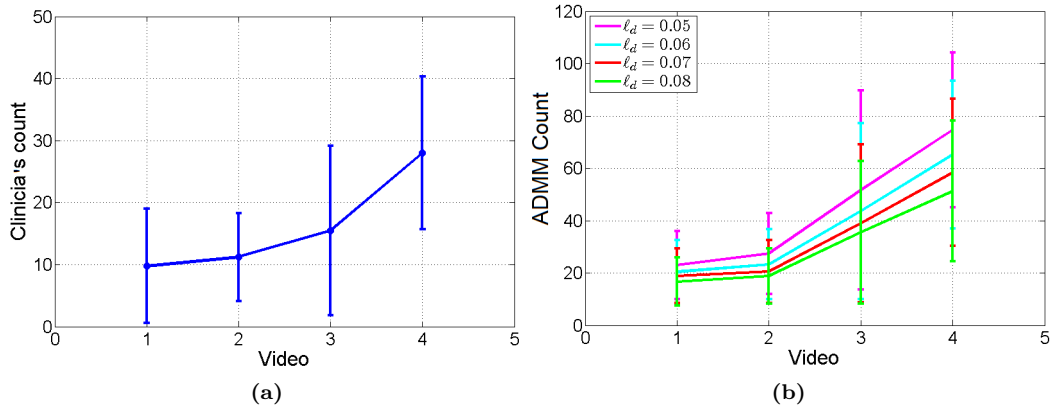


Figure 5.12 Mean number of detections per selected frames in videos 1 to 4 of Dataset II and the corresponding standard deviation. (a) clinician's opinion, (b) proposed method.

66.58% and precision of 23.64% : 32.4%, and provided the results in Fig. 5.12. We can observe that the counts of both the algorithm and the clinician increase as the bacteria concentration increases, which reflects the agreement between the approach considered and the clinician's annotations.

We can observe that the algorithm counts are higher than that of the clinician for the two processed datasets, as we expect the algorithm to be able to identify dots that are hard to be seen by the naked eye. Moreover, the clinician did not annotate ambiguous dots, meaning that a number of these were not chosen. This, along with false positives, is the main reason why the algorithm counts are higher than the clinician counts.

5.7.5 Comparison

In this subsection, we compare the proposed approach presented in this chapter with popular spot-detection methods from the literature, namely the Laplacian of Gaussian (LoG) and its approximation; the difference of Gaussians (DoG) filters [81, 97, 111], and the grey scale opening top-hat filter [33, 96]. These methods, although simple, have been considered in the literature of spot and blob detection in various applications. In this work, the LoG filter is implemented by employing a 5×5 kernel of standard deviation of 0.8 to each frame. Similarly, the DoG filter is implemented by considering the difference of two 5×5 Gaussian kernels of standard deviations of 0.5 and 0.8 respectively. On the other hand, the grey scale opening top-hat filter is employed by first smoothing the input image by a Gaussian kernel in order to reduce the noise, then by computing the morphological opening of the input image by employing a 3×3 flat disc, which achieves the best detection results and then subtracts the result from the original image. The same post processing steps described in the previous subsection are also employed in the resulting outlier amplitude images of these three existing methods, and the precision-recall curves are constructed accordingly. The comparison is conducted in terms of AUC of the resulting precision-recall curves, as well as in terms of computation time.

		Sparse coding based approach	LoG	DoG	Grey scale opening top-hat filter
	Videos	AUC			
Dataset I	1:3	**0.754**	0.58	0.56	0.749
	4:5	**0.8**	0.53	0.63	0.786
	6:7	**0.27**	0.175	0.104	0.172
	Average	**0.61**	0.43	0.43	0.569
Dataset II	1	**0.32**	0.142	0.14	0.257
	2	**0.43**	0.18	0.268	0.322
	3	**0.30**	0.09	0.116	0.184
	4	**0.26**	0.136	0.115	0.226
	Average	**0.33**	0.137	0.16	0.247

Table 5.3 Area under curve measures of the resulting precision-recall curves of the proposed approach and three existing methods.

	Sparse coding-based approach	LoG	DoG	Grey scale opening top-hat filter
Computation time	$5 \times \underline{0.4} = 2$	0.11	**0.05**	0.22

Table 5.4 Average computation time (in seconds) for the proposed method and three existing methods. In order to maintain a fair comparison between the four algorithms, the computational time of the sparse coding algorithm corresponds to the duration of five runs (used to select the best regularization parameter among the five values).

Table 5.3 compares the maximum achieved AUC of the proposed algorithm for Datasets I and II with those of the three methods described above. We can observe that the proposed algorithm provides the highest AUC for both Datasets I and II. Although the grey scale opening top-hat filter provides competitive results for videos 1:3 and 4:5 in Dataset I, it fails to identify the control cases as good as the proposed approach. The LoG and the DoG filters, on the other hand, showed similar performance.

On the other hand, Table 5.4 provides the average computation time of the four methods. As the LoG, DoG, and grey scale opening top-hat filters can provide results from just a single run, the proposed approach requires testing a number of regularization parameters to select the best performance in terms of bacteria detection. In a similar way to the way in which we computed the computation time for the deconvolution using the ADMM algorithm presented in Chapter 4, in this work, we consider computing the computation time of five runs of different outlier regularization parameter (β) values, when the sparse codes regularization parameter (α) is fixed. Although the proposed approach provides the highest computation time, it crucially brings the benefit of providing higher detection performance with respect to the other three methods.

5.8 Conclusions

In this chapter, we have illustrated the performance of a patch-based sparse representation approach for bacterial detection in OEM images of distal lung tissue using targeted SmartProbes.

We learned a dictionary for background image structure appearance (elastin, collagen, etc.) using a dictionary learning step, which was then used to predict any deviating outliers in testing frames. We formulated the bacteria detection problem with a sparse coding approach and provided a numerical solver in order to estimate the unknown model parameters. We have provided simulations on two ovine lung datasets instilled with bacteria, including a combination of fluorescent dyes, bacterial species and bacterial loads, which demonstrated that the estimated bacterial counts correlates with the bacterial counts performed by a clinician and good AUCs were achieved. The proposed algorithm was also compared with existing methods and showed superior performance in terms of bacteria detection. Although the advantages the proposed approach showed in terms of bacteria detection performance and reduced computation cost, it has some limitations. First, it requires the user to set regularization parameters that are difficult to tune when bacteria concentration varies in frames and/or videos. Second, it requires choosing appropriate cut-off threshold to the resulting outlier amplitude image in order to identify bacteria locations. The third limitation is related to the dictionary learning process. While annotating ground truth, it is highly likely that the annotator makes mistakes: (s)he can either falsely annotate a bacterium when it is noise, or simply miss-annotating a bacterium due to their overwhelming numbers in each frame. These types of error are common in any annotation process, but it might have a more severe impact on learning the dictionary since our target objects are 'dots' with similar structure. Therefore, wrongly annotated/un-annotated bacteria can provide biased dictionary atoms that cause errors in the estimation process. The next chapter overcomes these limitations and provides fully automatic unsupervised-based approach for the detection of fluorescently labelled bacteria using a hierarchical Bayesian model.

Chapter 6

Bayesian Bacterial Detection

This chapter has been adapted from the journal paper [Eldaly et al. (2018b)] (submitted).

Contents

6.1 Introduction

As previously mentioned in Chapters 3 and 5, in the literature, the object detection problem has been mainly addressed by either supervised or unsupervised approaches. In supervised approaches [13, 87, 108, 164], the dataset is usually divided into training, testing and validation sets. In the training phase, a model is trained by pairing inputs with their expected outputs which are also known as the ground truth. This trained model is then used to estimate the output of the test dataset. The validation dataset is often used to estimate model hyperparameters. While these algorithms are usually simple and fast, the trained model and its detection performance usually depend on the ground truth considered when training the model/classifier. Since the bacteria annotation task is subjective, meaning that it might be different from one clinician/biologist to another, therefore, considering such methods for cases where the ground truth can be highly subjective can result in a biased trained model. Moreover, some of these models use convolutional neural networks [13], which is challenging here due to the lack of sufficient training images and difficulty of bacteria annotations. On the other hand, some studies have considered unsupervised approaches, by using linear or morphological filtering methods [33, 96, 170, 171], or hierarchical Bayesian models [9, 53, 124], to solve outlier detection problems. These methods assume appearance model for the objects of interest. In particular, hierarchical Bayesian models can deal with uncertainty in inference due to the limitation in amount of data or prior information. Moreover, other unknown parameters can be jointly estimated within the algorithm such as outlier concentration(s), noise variance(s) and other model hyperparameters.

In this chapter, in order to overcome the limitations of the bacteria detection model presented in Chapter 5, and motivated by the models considered in [9, 53, 124], we propose a Bayesian approach for bacterial detection in OEM images. The model considered assumes that the observed pixel fluorescence is a linear combination of the actual intensity value associated with tissues or background, corrupted by additive Gaussian noise and potentially by an additional sparse outlier term modelling anomalies (which are considered to be candidate bacteria). The bacteria detection problem is formulated in a Bayesian framework and prior distributions are assigned to the unknown model parameters. A Markov chain Monte Carlo algorithm based on a partially collapsed Gibbs sampler is then used to sample the posterior distribution of the unknown parameters.

Contributions

The main contributions of this work are fourfold and can be summarized as below:

1. We develop an algorithm that can help to automatically detect labelled bacteria in optical endomicroscopy images. To the best of our knowledge, it is the first time this problem is addressed in a statistical framework by using a hierarchical Bayesian model. The proposed

model does not rely on a particular spatial organisation of the fibre bundle used in imaging the samples, hence it can be applied to various fibre bundle designs.

2. We propose a fully automatic algorithm in the sense that it allows the automated estimation of the model hyperparameters. Thus, it does not require the user to tune crucial parameters.

3. We provide simulations using synthetic and real datasets. For synthetic data, we model same system output and test different scenarios of outlier percentages, means and variances, and noise levels. For real datasets, we investigate different combinations of SmartProbes and bacterial concentrations including control cases for which no bacteria are present. Different SmartProbes that cause weak and stronger fluorescence are tested. We also show that the algorithm can differentiate controls from bacterial loads of different concentrations.

4. We compare the real data results of the proposed model with bacteria annotations performed by a trained clinician, and the sparse coding approach proposed in Chapter 5 in addition to the three existing methods presented in the same chapter, using both dot-annotation and count-annotation methods.

The remaining sections of the chapter are organized as follows. Section 6.2 formulates the problem of bacteria detection in OEM images, followed by Section 6.3 which summarizes the likelihood and the prior distributions assigned to the unknown parameters of the model. The resulting joint posterior distribution and the partially collapsed Gibbs sampler used to sample that distribution are discussed in Section 6.4. Simulations conducted on synthetic dataset that models same system output are provided in Section 6.5. Results and discussions on two real datasets of *ex vivo* ovine lungs instilled with bacteria are presented in Section 6.6. Conclusions and future work are finally reported in Section 6.7.

6.2 Formulation to the Bacteria Detection Problem Using a Bayesian Approach

As previously presented in Chapter 4, since only fibre cores hold information about the object being imaged, each image can be sparsely represented with only N_1 measurements, associated with N_1 fibres, where each fibre core is represented by a single intensity value. The size of OEM images in our case is 274×384 pixels and each of these images consists of $N_1 = 12105$ fibre cores. This corresponds to approximately 8.69% of the original image pixels. The interpolated images contain the same amount of acquired information as the sets of core intensities, and these images are mostly used for visualization purposes. However, processing only central core intensities, whose coordinates are known from factory pre-calibration, reduces the data volume to be processed and is expected to result in faster algorithms than when considering the whole interpolated images.

We consider a set of N_1 observed sub-sampled intensities $\mathbf{y} = [y_1, \ldots, y_{N_1}]^T$. In a similar manner to [9, 53, 136], each of these samples is assumed to result from a linear combination of an actual intensity value and additive Gaussian noise, potentially corrupted by additive outliers. The observation model can thus be expressed as

$$\mathbf{y} = \mathbf{x} + \mathbf{r} + \mathbf{w}, \tag{6.1}$$

with $\mathbf{x} = [x_1, \ldots, x_{N_1}]^T, \mathbf{r} = [r_1, \ldots, r_{N_1}]^T$ and $\mathbf{w} = [w_1, \ldots, w_{N_1}]^T$, where x_n is the actual intensity value of the nth sample, r_n represents a potential outlier (bacteria) and w_n represents the additive noise, which is assumed to be independently and identically distributed over the N_1 fibre cores. The noise is assumed to be Gaussian distributed with covariance matrix $\sigma^2\mathbf{I}_{N_1}$, denoted as $\mathbf{w} \sim \mathcal{N}(\mathbf{w}; \mathbf{0_{N_1}}, \sigma^2\mathbf{I_{N_1}})$, where $\sim$ reads "is distributed according to". This model can be seen as the same as the supervised-based model presented in Chapter 5 but with representing the learned background structures term, namely $\mathbf{X} = \mathbf{D\Psi}$ with smooth intensities $\mathbf{x}$.

The problem investigated in this work is to estimate the outlier vector $\mathbf{r}$ in Eq.(6.1), but the intensity values $\mathbf{x}$ and the noise variance σ^2 are also unknown. Thus we estimate jointly $\mathbf{r}, \mathbf{x}$ and σ^2 from the observation vector $\mathbf{y}$. To solve this problem, we propose a hierarchical Bayesian model and a sampling method to estimate the unknown parameters.

6.3 Hierarchical Bayesian Model

This section introduces the hierarchical Bayesian model proposed to estimate the unknown parameter $\mathbf{x}$, $\mathbf{r}$ and σ^2. This model is based on the likelihood function of the observations and on prior distributions assigned to the unknown parameters.

6.3.1 Likelihood

Eq. (6.1) implies that $\mathbf{y}|(\mathbf{x}, \mathbf{r}, \sigma^2) \sim \mathcal{N}(\mathbf{y}; \mathbf{x} + \mathbf{r}, \sigma^2\mathbf{I}_{N_1})$. Consequently, the likelihood function of $\mathbf{y}$ can be expressed as

$$f(\mathbf{y}|\mathbf{x}, \mathbf{r}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{N_1/2} \exp\left(-\frac{\|\mathbf{y} - \mathbf{x} - \mathbf{r}\|_2^2}{2\sigma^2}\right). \tag{6.2}$$

6.3.2 Parameter Prior Distributions

6.3.2.1 Intensity field x

For many applications, the intensity values of the scene to be recovered are likely to be spatially correlated. A classical and convenient way to model spatially correlated intensities is to consider Markov random fields (MRF) to build a prior model for $\mathbf{x}$ [59, 122, 157]. MRFs assume that the

distribution of a given intensity x_n, conditioned on the other intensity values of the image, reduces to its distribution conditioned on the values of its spatial neighbours, i.e., $f(x_n|\mathbf{x}_{\setminus x_n}) = f(x_n|\mathbf{x}_{\mathcal{V}_n})$, where $\mathcal{V}_n$ is the set of indices of the neighbours of x_n, $\mathbf{x}_{\setminus x_n}$ denotes the vector $\mathbf{x}$ whose element x_n has been removed, and $\mathbf{x}_{\mathcal{V}_n}$ is the subset of $\mathbf{x}$ composed of the elements whose indexes belong to $\mathcal{V}_n$. In this work, a Delaunay triangulation scheme is used on the N_1 samples to define the neighbourhood structure. Given the structure of the fibre bundle considered, each fibre core has between five and eleven neighbouring cores with mean distance between neighbours of 4.1 pixels. We specify $f(x_n|\mathbf{x}_{\mathcal{V}_n})$ as

$$f(x_n|\mathbf{x}_{\mathcal{V}_n}, \gamma_{\mathbf{x}}^2) \propto \exp\left(-\frac{1}{\gamma_{\mathbf{x}}^2} \sum_{n' \in \mathcal{V}_n} \frac{(x_n - x_{n'})^2}{d_{n,n'}}\right), \tag{6.3}$$

and where $\propto$ reads "is proportional to", $d_{n,n'}$ denotes the Euclidean distance between the spatial locations n and n', and the hyperparameter $\gamma_{\mathbf{x}}^2$ controls the global spatial correlation between intensities. Eq. (6.3) promotes smooth intensity variations between neighbours while ensuring that the prior dependence between neighbours decreases as $d_{n,n'}$ increases. The resulting joint prior $f(\mathbf{x}|\gamma_{\mathbf{x}}^2)$ can be expressed as [157]

$$f(\mathbf{x}|\gamma_{\mathbf{x}}^2) \propto \left(\gamma_{\mathbf{x}}^2\right)^{-(N_1-1)/2} \exp\left(-\frac{\mathbf{x}^T \boldsymbol{\Delta} \mathbf{x}}{2\gamma_{\mathbf{x}}^2}\right), \tag{6.4}$$

where

$$[\boldsymbol{\Delta}]_{i,j} = \begin{cases} \sum_{j \in \mathcal{V}_i} 1/d_{i,j} & \text{if } i = j, \\ 0 & \text{if} j \notin \mathcal{V}_i, \\ -1/d_{i,j} & \text{else.} \end{cases} \tag{6.5}$$

Note that each core has more that 5 neighbouring cores, and thus $\sum_{j \in \mathcal{V}_i} 1/d_{i,j} > 0, \forall i \in 1, \ldots, N_1$.

6.3.2.2 Noise variance σ^2

A Jeffreys' prior distribution [86], presented in Chapter 3, is chosen for the noise variance σ^2, i.e.,

$$f(\sigma^2) \propto \sigma^{-2} \mathbf{1}_{\mathbb{R}^+}(\sigma^2), \tag{6.6}$$

where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ denotes the indicator function defined on $\mathbb{R}^+$, which reflects the lack of knowledge about this parameter. This non-informative prior distribution can be easily replaced by conjugate inverse-Gamma prior distribution to include knowledge available about the noise level.

6.3.2.3 Outliers r

The concentration of bacteria in *in vivo* human lungs is expected to be such that only a small fraction of the fibres will be associated with bacteria detections. Hence the outliers are assumed

to be sparse, i.e., for most of the spatial locations, the outliers are expected to be exactly equal to zero. To model the outlier sparsity, we factorise the outlier vector as

$$\mathbf{r} = \mathbf{z} \odot \mathbf{t}, \tag{6.7}$$

where $\mathbf{z} = [z_1, \ldots, z_{N_1}]^T \in \{0,1\}^{N_1}$ is a binary label vector, $\mathbf{t} \in \mathbb{R}^{N_1}$ is the corresponding outlier amplitude vector, and $\odot$ denotes the Hadamard (term-wise) product. This decomposition allows one to decouple the location of the sparse components from their values. Precisely, $z_n = 1$ if a bacterium is present in the nth observed location with value equal to $r_n = t_n$, and $r_n = 0$ if there is no bacterium ($z_n = 0$). A conjugate Gaussian prior model is used for $\mathbf{t}$, i.e.,

$$f(\mathbf{t}|s_{\mathbf{t}}^2) = \prod_{n=1}^{N_1} \mathcal{N}\left(t_n; \mu_{\mathbf{t}}, s_{\mathbf{t}}^2\right), \tag{6.8}$$

where $\mu_{\mathbf{t}}$ and $s_{\mathbf{t}}^2$ control the prior mean and variance of the outliers, respectively. In this work, we don't assume a particular spatial structure for bacteria positions. However, they have a priori the same probability of being present in any region of the scene. To model this prior belief, we assign each label z_n the following Bernoulli prior distribution

$$f(z_n|\omega_{\mathbf{z}}) = \omega_{\mathbf{z}}\delta(1 - z_n) + (1 - \omega_{\mathbf{z}})\delta(z_n), \quad z_n \in \{0,1\}, \tag{6.9}$$

where $\delta(\cdot)$ denotes the Dirac delta function. Moreover, it is assumed that the probability of bacteria presence $\omega_{\mathbf{z}}$ is also unknown and we include this parameter within the inference process.

6.3.2.4 Hyperparameters

To reflect the lack of prior knowledge about the outlier variance in Eq. (6.8) and regularisation parameter $\gamma_{\mathbf{x}}^2$ in Eq.(6.4), the following weakly informative inverse-gamma prior distributions are assigned to $s_{\mathbf{t}}^2$ and $\gamma_{\mathbf{x}}^2$

$$\begin{cases} s_{\mathbf{t}}^2 \sim \mathcal{IG}(\eta, \nu), \\ \gamma_{\mathbf{x}}^2 \sim \mathcal{IG}(\eta, \nu), \end{cases} \tag{6.10}$$

where (η, ν) are fixed to $(\eta, \nu) = (10^{-3}, 10^{-3})$. Note that we did not observe significance change in the results when changing these hyperparameters. Similarly, we assign μ the following conjugate truncated Gaussian prior distribution

$$\mu_{\mathbf{t}}|(\bar{\mu}, \xi^2) \sim \mathcal{N}_{\mathbb{R}^+}\left(\mu_{\mathbf{t}}; \bar{\mu}, \xi^2\right), \tag{6.11}$$

where $(\bar{\mu}, \xi^2)$ are fixed and user-defined parameters (which depend on the dynamics of the image to be recovered). Truncated Gaussian prior distribution on the positive set is considered as we

expect outlier mean to be positive. In this work, we fixed $(\bar{\mu}, \xi^2) = (0, 10^6)$. Finally, we assign the bacteria presence percentage $\omega_{\mathbf{z}}$ a conjugate beta prior distribution

$$\omega_{\mathbf{z}} \sim \mathcal{B}e(\omega_{\mathbf{z}}; \alpha, \beta), \tag{6.12}$$

where (α, β) is fixed to $(\alpha, \beta) = (0.1, 1)$ as we expect the proportion of bacteria to be relatively small (the prior mean of $\omega_{\mathbf{z}}$ is $\alpha/(\alpha + \beta)$).

6.3.3 Joint Posterior Distribution

Assuming the parameters $\mathbf{x}, \mathbf{z}, \mathbf{t}$ and σ^2 are *a priori* mutually independent, the joint posterior distribution of the parameter vector $\boldsymbol{\Omega} = \{\mathbf{x}, \mathbf{z}, \mathbf{t}, \sigma^2\}$ and hyperparameters $\boldsymbol{\phi} = \{\mu_{\mathbf{t}}, s^2_{\mathbf{t}}, \gamma^2_{\mathbf{x}}, \omega_{\mathbf{z}}\}$ can be expressed as

$$f(\boldsymbol{\Omega}, \boldsymbol{\phi}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\Omega}) f(\boldsymbol{\Omega}|\boldsymbol{\phi}) f(\boldsymbol{\phi}), \tag{6.13}$$

where

$$f(\boldsymbol{\Omega}|\boldsymbol{\phi}) = f(\mathbf{x}|\gamma^2_{\mathbf{x}}) f(\mathbf{z}|\omega_{\mathbf{z}}) f(\mathbf{t}|\mu_{\mathbf{t}}, s^2_{\mathbf{t}}) f(\sigma^2), \tag{6.14}$$

$$f(\boldsymbol{\phi}) = f(\gamma^2_{\mathbf{x}}) f(\mu_{\mathbf{t}}) f(s^2_{\mathbf{t}}) f(\omega_{\mathbf{z}}), \tag{6.15}$$

and $f(\mathbf{z}|\omega_{\mathbf{z}}) = \prod_n f(z_n|\omega_{\mathbf{z}})$.

Fig. 6.1 depicts the directed acyclic graph (DAG) summarising the structure proposed Bayesian model.

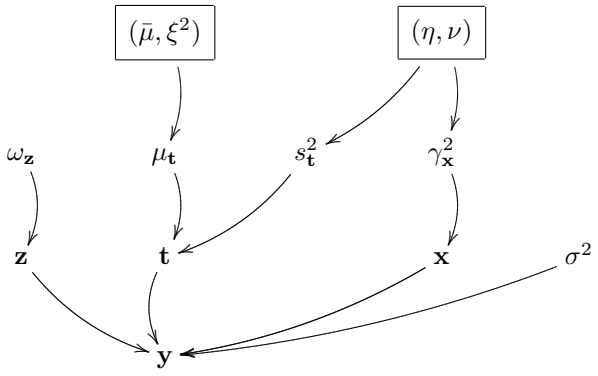


Figure 6.1 Graphical model for the proposed hierarchical Bayesian model (fixed quantities appear in boxes).

The next paragraph presents a sampling strategy to estimate the unknown parameter vector $\boldsymbol{\Omega}$ and the hyperparameters $\boldsymbol{\phi}$.

6.4 Bayesian Inference

To overcome the challenging derivation of Bayesian estimators associated with $f(\mathbf{\Omega}, \boldsymbol{\phi}|\mathbf{y})$, we propose to use an efficient Markov Chain Monte Carlo (MCMC) method to generate samples asymptotically distributed according to Eq.(6.13). In practice, strong correlations appear between $\mathbf{x}$ and $\mathbf{t}$, and between $\mathbf{z}$ and $\mathbf{t}$. Moreover, as $\mathbf{z}$ is very sparse, sampling $f(\mathbf{t}, \mu_{\mathbf{t}}, s^2_{\mathbf{t}}|\mathbf{y}, \mathbf{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi}_{\setminus(\mu_{\mathbf{t}}, s^2_{\mathbf{t}})})$, where $\mathcal{H}_{\setminus \mathbf{u}}$ denotes the parameter vector $\mathcal{H}$ whose $\mathbf{u}$ parameter is omitted, using a traditional Gibbs sampler results in very slow convergence. Hence, we propose a partially collapsed Gibbs sampler (PCGS) which yields better mixing and convergence properties of the generated Markov chain. A Gibbs sampler can be transformed into a partially collapsed Gibbs sampler by considering the following three basic tools. The first tool, marginalization, entails moving a group of unknowns from being conditioned upon to being sampled in one or more steps of a Gibbs sampler; the marginalized group can differ among the steps. Second, we may need to permute the steps of the sampler to use the third tool, trimming sampled components from the various steps that can be removed from the sampler without altering its Markov transition kernel. The PCGS used here samples groups of variables (e.g., $(\mathbf{x}, \mathbf{t})$) from their joint posterior distribution, in a similar fashion to block Gibbs samplers, which yields better mixing and convergence properties than sampling the variables (e.g., $\mathbf{x}$ and $\mathbf{t}$) sequentially from their conditional distributions. Sampling the joint distribution is achieved by first marginalising some variables which are then sampled from their full conditional distribution [113, 188]. Precisely, we propose to sample sequentially the elements of $\mathbf{\Omega}$ and $\boldsymbol{\phi}$ using moves that are summarised in Algorithm 12.

Algorithm 12 Partially Collapsed Gibbs Sampling Algorithm For Bacteria Detection - Version I

1: **Fixed input parameters**: Number of burn-in iterations N_{bi}, total number of iterations N_{MC}
2: **Initialization** $(k = 0)$
- Set $\mathbf{x}^{(0)}, \mathbf{z}^{(0)}, \mathbf{t}^{(0)}, {\sigma^2}^{(0)}, \gamma^{2\,(0)}_{\mathbf{x}}, \omega^{(0)}_{\mathbf{z}}, {\mu_{\mathbf{t}}}^{(0)}$

3: **Repeat** $(1 \leq k \leq N_{\mathbf{MC}})$
(a) Sample $\left(s^{2\,(k)}_{\mathbf{t}}, \mathbf{t}^{(k)}_0\right) | \left(\mathbf{y}, \mathbf{\Omega}^{(k-1)}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}^{(k-1)}_{\setminus s^2_{\mathbf{t}}}\right)$
(b) Sample $\left(\mu^{(k)}_{\mathbf{t}}, \mathbf{t}^{(k)}_0\right) | \left(\mathbf{y}, \mathbf{\Omega}^{(k-1)}_{\setminus \mathbf{t}_0}, \gamma^{2\,(k-1)}_{\mathbf{x}}, s^{2\,(k)}_{\mathbf{t}}, \omega^{(k-1)}_{\mathbf{z}}\right)$
(c) Sample $\gamma^{2\,(k)}_{\mathbf{x}} | \left(\mathbf{y}, \mathbf{\Omega}^{(k-1)}, \mu^{(k)}_{\mathbf{t}}, s^{2\,(k)}_{\mathbf{t}}, \omega^{(k-1)}_{\mathbf{z}}\right)$
(d) Sample $\omega^{(k)}_{\mathbf{z}} | \left(\mathbf{y}, \mathbf{\Omega}^{(k-1)}, \mu^{(k)}_{\mathbf{t}}, s^{2\,(k)}_{\mathbf{t}}, \gamma^{2\,(k)}_{\mathbf{x}}\right)$
(e) Sample $\left(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}\right) | \left(\mathbf{y}, \mathbf{z}^{(k-1)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)}\right)$
(f) Sample $\left(\mathbf{z}^{(k)}, \mathbf{t}^{(k)}\right) | \left(\mathbf{y}, \mathbf{x}^{(k)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)}\right)$
(g) Sample ${\sigma^2}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, \mathbf{t}^{(k)}, \mathbf{z}^{(k)}, \boldsymbol{\phi}^{(k)}\right)$

4: **Set** $k = k + 1$.

In Algorithm 12, $\mathbf{t}_0$ denotes the elements of $\mathbf{t}$ whose corresponding labels in $\mathbf{z}$ are null. Simi-

larly, $\mathbf{t}_1$ denotes the elements of $\mathbf{t}$ whose labels are equal to 1. We now detail each sampling step of Algorithm 12 as follow:

Sampling $(s^2_{\mathbf{t}}, \mathbf{t}_0)|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}$: As mentioned earlier, due to the sparsity of the outlier label vector $\mathbf{z}$, we propose to sample simultaneously $(s^2_{\mathbf{t}}, \mathbf{t}_0)|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}})$, which improves the mixing and convergence properties of the Markov chains, rather than considering a Gibbs sampler to sample from $s^2_{\mathbf{t}}|(\mathbf{y}, \boldsymbol{\Omega}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}})$ and $\mathbf{t}_0|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi})$. This is done by using

$$f\left(s^2_{\mathbf{t}}, \mathbf{t}_0|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}\right) = f\left(s^2_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}\right) f\left(\mathbf{t}_0|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}\right), \tag{6.16}$$

where

$$f\left(s^2_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}\right) = \int f\left(s^2_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}\right) d\mathbf{t}_0. \tag{6.17}$$

In other words, sampling from $(s^2_{\mathbf{t}}, \mathbf{t}_0)|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}})$ can be achieved by sampling sequentially $s^2_{\mathbf{t}}$ from $f\left(s^2_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}\right)$, then $\mathbf{t}_0$ from $f\left(\mathbf{t}_0|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}\right)$. It is easy to show that sampling from $f\left(s^2_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}\right)$ can be achieved by sampling from the following inverse-gamma distribution

$$s^2_{\mathbf{t}}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus s^2_{\mathbf{t}}}) \sim \mathcal{IG}\left(\eta + \frac{N'_1}{2}, \nu + \frac{\sum_{n \in I_1} (t_n - \mu_{\mathbf{t}})^2}{2}\right), \tag{6.18}$$

where $N'_1 = \text{card}(\mathbf{t}_1)$ and $I_{\mathbf{1}} = \{n|z_n = 1\}$.

On the other hand, when $z_n = 0$, t_n does not appear in Eq.(6.1). Thus sampling $\mathbf{t}_0|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi})$ reduces to sampling its elements independently from their Gaussian prior distribution defined in Eq.(6.8).

Sampling $(\mu_{\mathbf{t}}, \mathbf{t}_0)|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}})$: Similar to the outlier variance, we consider a PCGS to sample simultaneously $(\mu_{\mathbf{t}}, \mathbf{t}_0)|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}})$. This is done by using

$$f\left(\mu_{\mathbf{t}}, \mathbf{t}_0|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}}\right) = f\left(\mu_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}}\right) f\left(\mathbf{t}_0|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}\right), \tag{6.19}$$

where

$$f\left(\mu_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}}\right) = \int f\left(\mu_{\mathbf{t}}|\mathbf{y}, \boldsymbol{\Omega}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}}\right) d\mathbf{t}_0. \tag{6.20}$$

It is easy to show that sampling $\mu_{\mathbf{t}}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}})$ reduces to sampling from the following truncated Gaussian distribution

$$\mu_{\mathbf{t}}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi}_{\setminus \mu_{\mathbf{t}}}) \sim \mathcal{N}_{\mathbb{R}^+}(\mu_{\mathbf{t}}; \mathcal{M}, \mathcal{S}), \tag{6.21}$$

where

$$\mathcal{M} = \frac{\bar{\mu}s^2 + \xi^2 \sum_{n\in I_1} t_n}{\xi^2 N_1' + s_{\mathbf{t}}^2}, \quad \mathcal{S} = \frac{s_{\mathbf{t}}^2 \xi^2}{\xi^2 N_1' + s_{\mathbf{t}}^2}. \tag{6.22}$$

Sampling from Eq.(6.21) can be achieved efficiently by using the method proposed in [123] while sampling $\mathbf{t}_0|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}, \boldsymbol{\phi})$ reduces to sampling from Gaussian distributions, as discussed above.

Sampling $\gamma_{\mathbf{x}}^2|(\mathbf{y}, \boldsymbol{\Omega}, \boldsymbol{\phi}_{\setminus \gamma_{\mathbf{x}}^2})$: By cancelling out the terms that do not depend on $\gamma_{\mathbf{x}}^2$ from the posterior distribution in Eq.(6.13), its conditional distribution reduces to the following inverse-gamma distribution

$$\gamma_{\mathbf{x}}^2|(\mathbf{y}, \boldsymbol{\Omega}, \boldsymbol{\phi}_{\setminus \gamma_{\mathbf{x}}^2}) \sim \mathcal{IG}\left(\eta + \frac{N_1 - 1}{2}, \nu + \frac{\mathbf{x}^T \boldsymbol{\Delta} \mathbf{x}}{2}\right), \tag{6.23}$$

which is easy to sample from.

Sampling $\omega_{\mathbf{z}}|(\mathbf{y}, \boldsymbol{\Omega}, \boldsymbol{\phi}_{\setminus \omega_{\mathbf{z}}})$: Due to the conjugacy of the hierarchical prior model $f(\mathbf{z}|\omega_{\mathbf{z}})f(\omega_{\mathbf{z}})$, the full conditional distribution of $\omega_{\mathbf{z}}$ reduces to the following beta distribution

$$\omega_{\mathbf{z}}|(\mathbf{y}, \boldsymbol{\Omega}, \boldsymbol{\phi}_{\setminus \omega_{\mathbf{z}}}) \sim \mathcal{B}e\left(\alpha + \sum_{n=1}^{N_1} z_n, \beta + N_1 - \sum_{n=1}^{N_1} z_n\right). \tag{6.24}$$

Sampling $(\mathbf{x}, \mathbf{t})|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{x},\mathbf{t})}, \boldsymbol{\phi})$: As mentioned earlier, we propose to sample simultaneously $(\mathbf{x}, \mathbf{t})|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{x},\mathbf{t})}, \boldsymbol{\phi})$, rather than considering a Gibbs sampler to sample from $\mathbf{x}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{x}}, \boldsymbol{\phi})$ and $\mathbf{t}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi})$. This is done by using

$$f\left(\mathbf{x}, \mathbf{t}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{x},\mathbf{t})}, \boldsymbol{\phi}\right) = f\left(\mathbf{x}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{x},\mathbf{t})}, \boldsymbol{\phi}\right) f\left(\mathbf{t}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi}\right), \tag{6.25}$$

where

$$f\left(\mathbf{x}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{x},\mathbf{t})}, \boldsymbol{\phi}\right) = \int f\left(\mathbf{x}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{x}}, \boldsymbol{\phi}\right) d\mathbf{t}. \tag{6.26}$$

It is easy to show that $f\left(\mathbf{x}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{x},\mathbf{t})}\right)$ is the following multivariate Gaussian distribution

$$\mathbf{x}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{x},\mathbf{t})}, \boldsymbol{\phi}) \sim \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{6.27}$$

where

$$\begin{cases} \boldsymbol{\mu} = \boldsymbol{\Sigma}(s_{\mathbf{t}}^2 \mathbf{Z} + \sigma^2 \mathbf{I})^{-1}(\mathbf{y} - \mu_{\mathbf{t}}\mathbf{z}), \\ \boldsymbol{\Sigma} = \left((s_{\mathbf{t}}^2 \mathbf{Z} + \sigma^2 \mathbf{I})^{-1} + \gamma^{-2}\boldsymbol{\Delta}\right)^{-1}, \end{cases} \tag{6.28}$$

and $\mathbf{Z}$ is a diagonal matrix of outlier labels. The full conditional distribution of $\mathbf{t}$, i.e., $f\left(\mathbf{t}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi})\right)$

reduces to the following multivariate Gaussian distribution

$$\mathbf{t}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi}) \sim \mathcal{N}(\mathbf{t}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{6.29}$$

where

$$\begin{cases} \boldsymbol{\mu} = \left(\frac{(\mathbf{y}-\mathbf{x})^T \odot \mathbf{z}}{\sigma^2} + \frac{\mu_t}{s^2} \right) \boldsymbol{\Sigma}, \\ \boldsymbol{\Sigma} = \left(s^{-2}\mathbf{I}_N + \sigma^{-2}\mathbf{Z} \right)^{-1}. \end{cases} \tag{6.30}$$

Note that $\boldsymbol{\Sigma}$ is a diagonal covariance matrix, which is easy to construct.

Sampling $(\mathbf{z}, \mathbf{t})|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{z},\mathbf{t})}, \boldsymbol{\phi})$: Updating simultaneously $(\mathbf{z}, \mathbf{t})$ is achieved using

$$f\left(\mathbf{z}, \mathbf{t}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{z},\mathbf{t})}, \boldsymbol{\phi}\right) = f\left(\mathbf{z}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{z},\mathbf{t})}, \boldsymbol{\phi}\right) f\left(\mathbf{t}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi}\right). \tag{6.31}$$

It can be seen from Eq.(6.13) that

$$f(z_n = m|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(z_{n'}, t_{n'})}, \boldsymbol{\phi}) \propto \bar{\omega_{\mathbf{z}}}_n^{(m)} \quad \forall(n), \tag{6.32}$$

where $m \in \{0, 1\}$ and

$$\log(\bar{\omega_{\mathbf{z}}}_n^{(m)}) = -\frac{1}{2}\log\left(2\pi(\sigma^2 + m s_{\mathbf{t}}^2)\right) - \frac{(y_n - x_n - m\mu_{\mathbf{t}})^2}{2(\sigma^2 + m s_{\mathbf{t}}^2)} + \log(p(m|\omega_{\mathbf{z}})).$$

Consequently, the label z_n can be drawn from its conditional distribution (where t_n has been marginalised) by drawing randomly from $\{0, 1\}$ with probabilities given by

$$f(z_n = m|\mathbf{y}, \boldsymbol{\Omega}_{\setminus(z_{n'}, t_{n'})}, \boldsymbol{\phi}) = \frac{\bar{\omega_{\mathbf{z}}}_n^{(m)}}{\bar{\omega_{\mathbf{z}}}_n^{(0)} + \bar{\omega_{\mathbf{z}}}_n^{(1)}}. \tag{6.33}$$

Moreover, the elements of $\mathbf{z}$ can be updated in a parallel manner using the fact that $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi}) = \prod_n f(z_n|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi})$. Sampling from $f(\mathbf{t}|\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi})$ can then be achieved using Eq. (6.29).

Sampling $\sigma^2|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \sigma^2}, \boldsymbol{\phi})$: In a similar fashion to the regularisation parameter $\gamma_{\mathbf{x}}^2$, the noise variance σ^2 can be sampled from the following inverse-gamma distribution

$$\sigma^2|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \sigma^2}, \boldsymbol{\phi}) \sim \mathcal{IG}\left(\frac{N_1}{2}, \frac{\|\mathbf{y} - \mathbf{x} - \mathbf{r}\|_2^2}{2}\right), \tag{6.34}$$

which is easy to sample from.

Algorithm 13 is a detailed version of Algorithm 12 following the sampling steps explained above. Although Algorithm 13 seems more complex (more sequential steps) than Algorithm 12, several steps can be omitted since some generated variables are not actually used. For instance,

Algorithm 13 Partially Collapsed Gibbs Sampling Algorithm to Bacteria Detection for OEM - Version II

1: **Fixed input parameters**: Number of burn-in iterations N_{bi}, total number of iterations N_{MC}

2: **Initialization** $(k = 0)$

- Set $\mathbf{x}^{(0)}, \mathbf{z}^{(0)}, \mathbf{t}^{(0)}, {\sigma^2}^{(0)}, {\gamma_{\mathbf{x}}^2}^{(0)}, \omega_{\mathbf{z}}^{(0)}, {\mu_{\mathbf{t}}}^{(0)}$

3: **Repeat** $(1 \leq k \leq N_{\mathbf{MC}})$

(a1) Sample $s_{\mathbf{t}}^{2(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}^{(k-1)}, \boldsymbol{\phi}_{\setminus s_{\mathbf{t}}^2}^{(k-1)} \right)$ from Eq.(6.18).

(a2) Sample $\mathbf{t}_0^{(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}^{(k-1)}, s_{\mathbf{t}}^{2(k)}, \mu_{\mathbf{t}}^{(k-1)}, \gamma_{\mathbf{x}}^{2(k-1)}, \omega_{\mathbf{z}}^{(k-1)} \right)$ from Eq.(6.8).

(b1) Sample $\mu_{\mathbf{t}}^{(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}^{(k-1)}, \gamma_{\mathbf{x}}^{2(k-1)}, s_{\mathbf{t}}^{2(k)}, \omega_{\mathbf{z}}^{(k-1)} \right)$ from Eq.(6.21).

(b2) Sample $\mathbf{t}_0^{(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}^{(k-1)}, \mu_{\mathbf{t}}^{(k)}, \gamma_{\mathbf{x}}^{2(k-1)}, s_{\mathbf{t}}^{2(k)}, \omega_{\mathbf{z}}^{(k-1)} \right)$ from Eq.(6.8).

(c) Sample $\gamma_{\mathbf{x}}^{2(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}^{(k-1)}, \mu_{\mathbf{t}}^{(k)}, s_{\mathbf{t}}^{2(k)}, \omega_{\mathbf{z}}^{(k-1)} \right)$ from Eq.(6.23).

(d) Sample $\omega_{\mathbf{z}}^{(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}^{(k-1)}, \mu_{\mathbf{t}}^{(k)}, s_{\mathbf{t}}^{2(k)}, \gamma_{\mathbf{x}}^{2(k)} \right)$ from Eq.(6.24).

(e1) Sample $\mathbf{x}^{(k)} | \left(\mathbf{y}, \mathbf{z}^{(k-1)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)} \right)$ from Eq.(6.27).

(e2) Sample $\mathbf{t}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, \mathbf{z}^{(k-1)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)} \right)$ from Eq.(6.29).

(f1) Sample $\mathbf{z}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)} \right)$ from Eq.(6.33).

(f2) Sample $\mathbf{t}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, \mathbf{z}^{(k)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)} \right)$ from Eq.(6.29).

(g) Sample ${\sigma^2}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, \mathbf{t}^{(k)}, \mathbf{z}^{(k)}, \boldsymbol{\phi}^{(k)} \right)$ from Eq.(6.34).

4: **Set** $k = k + 1$.

the variables generated in steps (a2) and (b2) are not used during the following steps and the steps (a2) and (b2) can thus be omitted. Similarly, sampling $\mathbf{t}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}}, \boldsymbol{\phi})$ in (e2) is omitted as this step is not required when sampling $\mathbf{z}|(\mathbf{y}, \boldsymbol{\Omega}_{\setminus(\mathbf{z},\mathbf{t})}, \boldsymbol{\phi})$ in (f1). These simplifications result in the final algorithm shown in Algorithm 14.

The algorithm is stopped after N_{MC} iterations, including N_{bi} burn-in iterations which correspond

Algorithm 14 Partially Collapsed Gibbs Sampling Algorithm to Bacteria Detection for OEM - Final Version

1: **Fixed input parameters**: Number of burn-in iterations N_{bi}, total number of iterations N_{MC}
2: **Initialization** $(k = 0)$
- Set $\mathbf{x}^{(0)}, \mathbf{z}^{(0)}, \mathbf{t}^{(0)}, {\sigma^2}^{(0)}, {\gamma_{\mathbf{x}}^2}^{(0)}, \omega_{\mathbf{z}}^{(0)}, {\mu_{\mathbf{t}}}^{(0)}$

3: **Repeat** $(1 \leq k \leq N_{\mathrm{MC}})$

(a1) Sample $s_{\mathbf{t}}^{2(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}^{(k-1)}, \boldsymbol{\phi}_{\setminus s_{\mathbf{t}}^2}^{(k-1)}\right)$ from Eq.(6.18).

(b1) Sample $\mu_{\mathbf{t}}^{(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}_{\setminus \mathbf{t}_0}^{(k-1)}, {\gamma_{\mathbf{x}}^2}^{(k-1)}, s_{\mathbf{t}}^{2(k)}, \omega_{\mathbf{z}}^{(k-1)}\right)$ from Eq.(6.21).

(c) Sample ${\gamma_{\mathbf{x}}^2}^{(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}^{(k-1)}, \mu_{\mathbf{t}}^{(k)}, s_{\mathbf{t}}^{2(k)}, \omega_{\mathbf{z}}^{(k-1)}\right)$ from Eq.(6.23).

(d) Sample $\omega_{\mathbf{z}}^{(k)} | \left(\mathbf{y}, \boldsymbol{\Omega}^{(k-1)}, \mu_{\mathbf{t}}^{(k)}, s_{\mathbf{t}}^{2(k)}, {\gamma_{\mathbf{x}}^2}^{(k)}\right)$ from Eq.(6.24).

(e1) Sample $\mathbf{x}^{(k)} | \left(\mathbf{y}, \mathbf{z}^{(k-1)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)}\right)$ from Eq.(6.27).

(f1) Sample $\mathbf{z}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)}\right)$ from Eq.(6.33).

(f2) Sample $\mathbf{t}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, \mathbf{z}^{(k)}, {\sigma^2}^{(k-1)}, \boldsymbol{\phi}^{(k)}\right)$ from Eq.(6.29).

(g) Sample ${\sigma^2}^{(k)} | \left(\mathbf{y}, \mathbf{x}^{(k)}, \mathbf{t}^{(k)}, \mathbf{z}^{(k)}, \boldsymbol{\phi}^{(k)}\right)$ from Eq.(6.34).

4: **Set** $k = k + 1$.

to the transient period of the sampler (determined visually from preliminary runs). The first N_{bi} samples are discarded and the remaining samples are used to approximate the following estimators. The label vector $\mathbf{z}$ is estimated using marginal maximum a posteriori (MAP) estimation. This estimator is then used to compute the minimum mean square error (MMSE) of $\mathbf{r}$ conditioned on $\mathbf{z} = \hat{\mathbf{z}}_{\mathrm{MAP}}$, i.e., $\hat{\mathbf{r}} = (\hat{\mathbf{r}}_{\mathrm{MMSE}}|\widehat{\mathbf{z}}_{\mathrm{MAP}}) \odot \widehat{\mathbf{z}}_{\mathrm{MAP}}$. Finally, the remaining parameters are estimated using the empirical averages of the generated samples (MMSE estimates), for instance, the MMSE of the actual intensity vector $\mathbf{x}$, denoted as $\hat{\mathbf{x}}$, is given by

$$\hat{\mathbf{x}} = \frac{1}{N_{\mathrm{MC}} - N_{\mathrm{bi}}} \sum_{t=N_{\mathrm{bi}+1}}^{N_{\mathrm{MC}}} \mathbf{x}^{(t)}. \tag{6.35}$$

(a) (b) (c)

Figure 6.2 Creation of the synthetic data: (a) Original image, (b) Natural neighbour interpolation from roughly 1.4% samples of the image in a, (c) corruption by noise ($\sigma^2 = 10$) and 5% of outliers $(\mu, s_{\mathbf{t}}^2) = (255, 100)$.

6.5 Synthetic Data Experiments

6.5.1 Data creation

In order to assess the proposed approach for denoising and detecting fluorescently labelled bacteria, the standard 'Lena' test image (256×256 pixels) is used. A subsampled version of this image is obtained by considering the sampling pattern of an actual OEM system, as mentioned in Chapter 4, which yields 928 randomly subsampled pixels (1.4% of the original image).

Fig. 6.2 shows the original Lena image, natural neighbour interpolation [166, Chap. 2] of the 928 randomly subsampled pixels, and an example of system output after applying the model in equation Eq.(6.1). In this case, noise variance ($\sigma^2 = 10$) and 5% of outliers with mean ($\mu_{\mathbf{t}} = 255$), and variance ($s_{\mathbf{t}}^2 = 100$) are used.

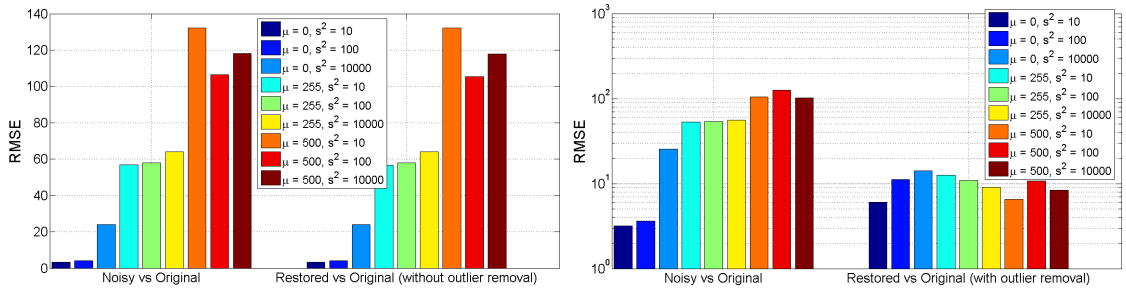
6.5.2 Evaluation criterion

For synthetic data, since the ground truth is available, we evaluate the algorithm not only on outlier detection, but also on denoising of fibre core intensities. The performance discriminator adopted to measure the quality of the denoised fiber cores is the root mean square error (RMSE), which is computed using intensities at the core locations using

$$\mathrm{RMSE}(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\frac{\sum_{n=1}^{N_1} \left(\mathbf{x}(n) - \hat{\mathbf{x}}(n)\right)^2}{N_1}}, \tag{6.36}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the subsampled reference Lena image and its denoised version respectively. The results are compared with median filtering in which a window size of 15×15 is used.

On the other hand, in order to assess the outlier detection performance, considering the fibre



(a) **(b)**

Figure 6.3 A plot of RMSE before and after denoising for $\sigma^2 = 10$ and 5% outliers, (a) without and (b) with outlier removal implementation.

cores that are neither annotated by the clinician nor detected by the proposed algorithm as true negatives would result in a heavily unbalanced two classes problem, hence, empirical precision-recall measures is considered, which are computed in the same way presented in Chapter 5.

6.5.3 Performance analysis

In order to assess the performance of the proposed method, weak and strong outliers and noise are tested and the RMSE is evaluated accordingly. We evaluated the algorithm by selecting $\sigma^2 \in \{0, 1, 10\}$, $\mu_{\mathbf{t}} \in \{0, 255, 500\}$, $s_{\mathbf{t}}^2 \in \{10, 100, 10000\}$ and outlier percentage of 1%, 5% and 10%. We show the results for $\sigma^2 = 10$, and outlier percentages of 5%. Results for outlier percentages of 1% and 10% are not shown here since the trends in behaviour are the same as 5%.

We consider two cases, either denoising without outlier removal (e.g., setting $\mathbf{z} = \mathbf{0}$ in Eq. (6.1)) and with outlier removal. Fig. 6.3 shows plots of RMSEs before and after denoising versus different outlier means and variances for denoising without (left) and with (right) outlier removal. In case of denoising without outlier removal, we can observe that there is no enhancement since RMSE before and after denoising are approximately equal. On the other hand, if outlier removal is implemented, RMSE after denoising is much lower than that before. However, if outlier mean and variance are similar to those of the added Gaussian noise, then the RMSE is higher, as the algorithm can not distinguish between the outliers and the noise in this case.

Median filtering can be used to mitigate, but not explicitly detect outliers, and hence is used for comparison purposes in terms of denoising. Fig. 6.4 shows a plot of RMSE versus different outlier means and variances for both the proposed approach and median filtering. We can observe that the proposed algorithm outperforms the median filter in terms of RMSE after denoising.

The precision-recall measures of the tested scenarios are represented by the empirical precision-recall graph shown in Fig. 6.5. We can observe the detection accuracy for all of the tests except for the cases where outlier mean and/or variance is near to noise variance, in which the outliers are of low amplitude, and hence could not be detected. This should not be a problem as in real world scenarios, including the one investigated in this work, outlier mean and variance are positive

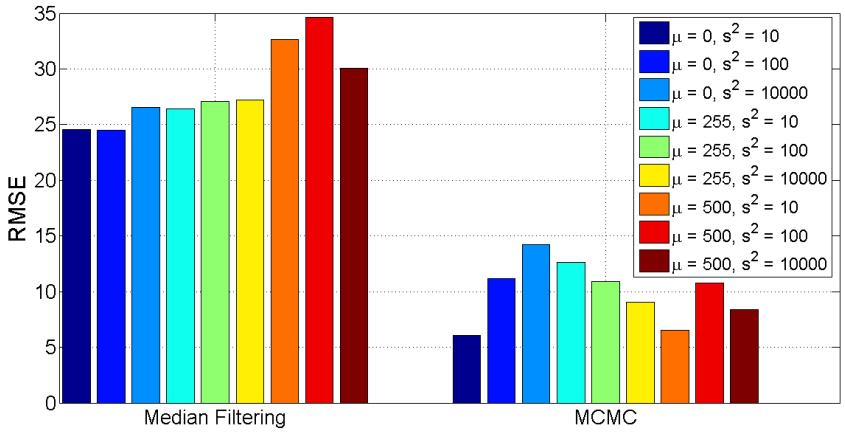


Figure 6.4 A comparison between median filtering and the proposed method.

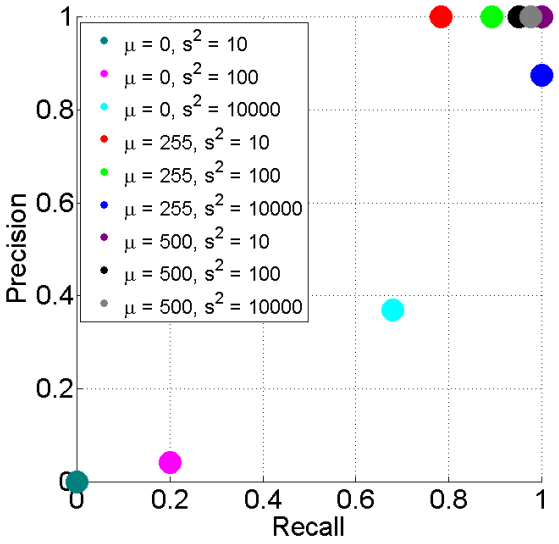
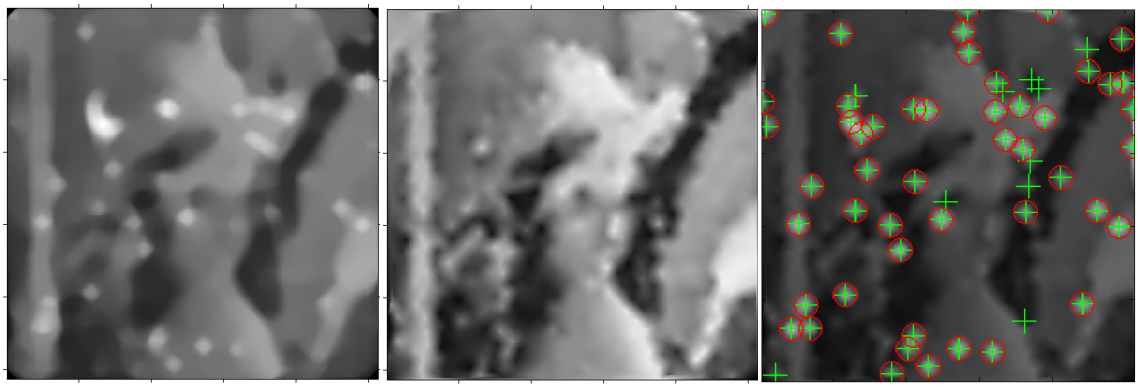


Figure 6.5 Precision-recall plot for tested outlier means and variances for $\sigma^2 = 10$, and outlier proportion of 5%.

and much higher than that of the noise.

Fig. 6.6(a) and Fig. 6.6(b) show the results of denoising of the image in Fig. 6.2(c) using both median filtering and the proposed approach respectively. We can observe that the proposed algorithm gives better result since the structure of the Lena image is more identified and less blurred. Fig. 6.6(c) shows the result of outlier detection. We can observe that all of the outliers are detected, however there are a few false negative detections which correspond to sharp intensity changes.



(a) (b) (c)

Figure 6.6 (a) Median filtering denoising, (b) MCMC denoising, and (c) result of outlier detection using MCMC, red circles represent true outlier locations whereas green pluses represent detections using the proposed approach.

6.6 Simulations Using Real Datasets

6.6.1 Datasets and algorithm evaluation

Datasets

The proposed algorithm is tested on the same two datasets described in Chapter 5. However, in order to adapt the datasets to current problem formulation, the locations of the clinician annotations are rounded to the closest core location. The fibre core locations are provided by the imaging system with sub-pixel accuracy, so the fibre core centre coordinates were also rounded to the nearest pixel.

Algorithm evaluation

The algorithm was run for each of the annotated frames, namely the 133 frames in Dataset I, and the 58 frames in Dataset II, yielding sets of detected bacteria, i.e potential labelled bacteria. Each bacterium is anticipated to be visible in a single core at a time. Hence, each group of connected detections is grouped and counted as a single detection. Once these connected detections are identified, they are replaced by a single detection at the mean of locations of these connected detections. This, along with the solo detections which do not have detections at their neighbouring cores gives the total number of detected bacteria in each frame.

Similar to the sparse coding approach presented in the previous chapter, we compare different criteria for evaluating the performance of the MCMC algorithm. Firstly, a statistical comparison of bacterial counts performed by the trained clinician and the algorithm output is considered (count-annotation). Secondly, in order to assess the bacteria detection performance, empirical precision-recall measures are considered. Note that this algorithm provides directly the locations of detected

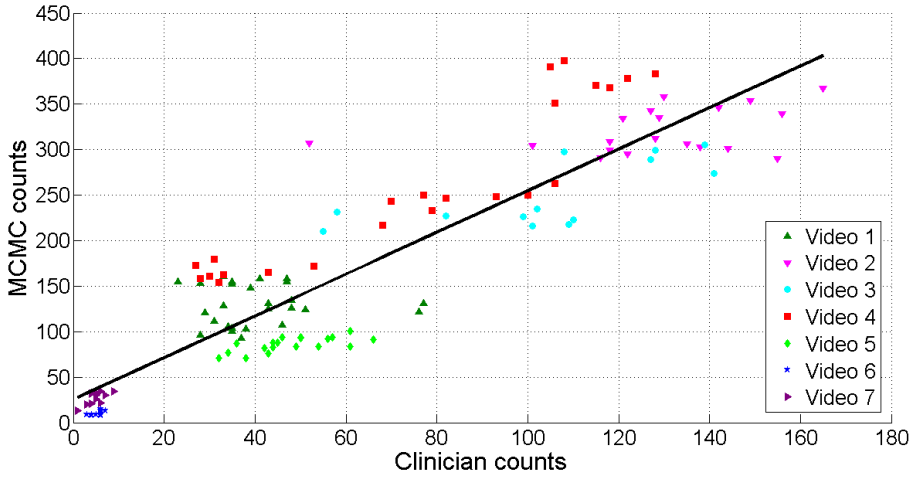


Figure 6.7 Plot of clinician bacteria count versus algorithm bacteria count for dataset I. Dots correspond to frames, and colours correspond to videos.

outliers which are defined by the label vector $\mathbf{z}$. Hence, in the contrary to the sparse coding algorithm presented in Chapter 5 where the outlier amplitude image needed to be thresholded (by different ℓ_d) and a precision-recall curve was plotted accordingly, only single precision-recall value is available for each frame, in this approach.

Figure 6.7 provides a scatter plot of the number of labelled anomalies (potential bacteria) identified manually and automatically in each frame of each video in Dataset I. A range of bacteria and fluorophore dyes are considered, on subjects with fixed bacterial concentration as well as controls (including fluorophore but no bacteria). Similarly, for Dataset II, Figure 6.8 depict a plot of the total bacteria counts performed by the clinician versus that by the algorithm for different bacterial concentrations. Table 6.1 provides precision-recall measures for both Datasets I and II.

6.6.2 Discussion

Count-annotation effect

For Dataset I, similar to the unsupervised-based bacteria detection algorithm presented in the previous chapter, the algorithm counts are compared with the clinician counts in each frame as shown in Fig. 6.7. We can observe an almost linear relationship between the clinician counts and algorithm counts. Indeed the empirical correlation between the manually and automatically detected anomalies is 0.912.

Moreover, for videos 1, 2 and 3 in which a highly fluorescent SmartProbe is used, and videos 4 and 5 in which an in-house SmartProbe which produces weaker fluorescence signal is used, a similar trend is observed between the clinician and the algorithm. This also applies on the type of bacteria the samples are labelled with. Videos 6 and 7 which are controls, demonstrate minimal counts using both the clinician and the algorithm, which reflects the ability of the algorithm to

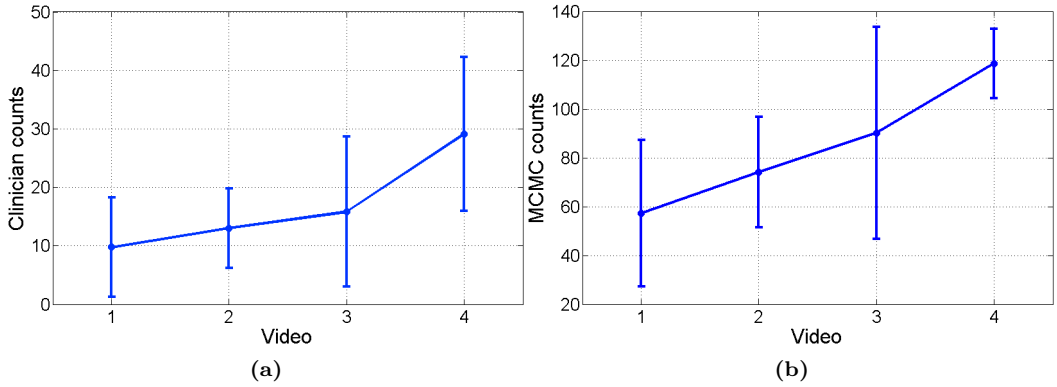


Figure 6.8 Mean number of detections per selected frames in videos 1 to 4 of Dataset II and the corresponding standard deviation. (a) clinician's opinion, (b) proposed method.

differentiate bacterial loads from control. For Dataset II, in Fig. 6.8, we can observe that both of the clinician and the algorithm counts increase with the bacteria concentration which reflects the agreement between the approach considered and the clinician's annotations.

In the two datasets, we can observe that the algorithm counts are higher than that of the clinician, as we expect the algorithm to be able to identify dots that are hard to be seen by the naked eye. Moreover, the clinician did not annotate ambiguous dots that a number of cases were not chosen. This, along with false positives are the main reasons why the algorithm counts are higher than the clinician counts.

Dot-annotation effect

In terms of the detection accuracy, Table. 6.1 provides precision-recall measures for different SmartProbe combinations in Dataset I and different bacterial concentration in Dataset II. We can observe the detection accuracy of the proposed approach due to the high values of the average recall measures, with reasonable average precision measures for the two processed datasets. This shows the benefit of having fully automatic and adaptive algorithm that lets the data speaks about itself, which leaded to infer the bacteria concentration in each individual frame. Similar to the sparse coding approach presented in the previous chapter, the algorithm provides slightly better precision-recall measures for videos 4:5 in which weaker SmartProbe was used to stain the samples compared to videos 1:3 in which a higher fluorescent SmartProbe is used. This in fact is due to the high amount of elastin structures present in videos 1:3 compared to videos 4:5, which increased the chance of having higher false positive detections. Though, the algorithm provides high detection performance which reflects its ability to detect bacteria in both cases.

For completeness, Fig. 6.9 shows examples of detections in two frames from Dataset I. Fig. 6.9 (a) and (b) show the original frames, (c) and (d) show the clinician's annotations along with all the algorithms detections superimposed, and finally (g) and (h) show the final algorithm detections

	Videos	Precision	Recall
Dataset I	1:3	35.36%	95.52%
	4:5	44.97%	99.00%
	6:7	23.14%	89.60%
	Average	37.94%	96.63%
Dataset II	1	15.50%	86.86%
	2	18.82%	98.35%
	3	17.27%	91.14%
	4	21.63%	82.61%
	Average	18.3%	89.74%

Table 6.1 Quantitative detection measures for datasets I and II.

after considering the post processing steps described in Section 6.6.1. We can observe the accuracy of the algorithm as almost all of the cores annotated by the clinician are covered by the algorithm's detections.

6.6.3 Comparison

In this subsection, we compare the results of the proposed approach presented in this chapter with the results of the sparse coding approach and the three existing spot/blob-detection methods presented in the previous chapter (Chapter 5). The comparison is conducted in terms of precision-recall measures and computation time.

Due to the absence of AUC measures for the Bayesian approach presented in this chapter, we consider comparing the four methods by finding the precision measures corresponding to the recall measures of the Bayesian approach. The comparison is presented in Table 6.2. It is clear that the performance of the Bayesian approach outperforms the rest of the methods for both Datasets I and II. Moreover, it is far superior in providing the highest precisions for the samples stained with weak SmartProbes (videos 4:5) and the control samples (videos 6:7). On the other hand, the sparse coding approach provides the second best results for Dataset II and close results with the grey scale top-hat filter for Dataset I. However, for this case, there might be a different precision-recall combination at which the sparse coding approach outperforms the grey scale top-hat filter.

On the other hand, Table 6.3 provides the average computation time of the five methods. As the Bayesian approach presented in this chapter can provide results from just a single run, the rest of the methods require thresholding the outlier amplitude image, which is hard to represent here. The computation time of the sparse coding approach corresponds to the duration of five runs of different outlier regularization parameter (β) values, when the sparse codes regularization parameter (α) is fixed. Although that the Bayesian proposed approach provides the highest computation time, it crucially brings the benefit of providing higher detection performance with respect to the other four methods, and being fully automatic, as there is no need to either set any regularization parameters or threshold the resulting outlier amplitude image in order to identify outlier locations.

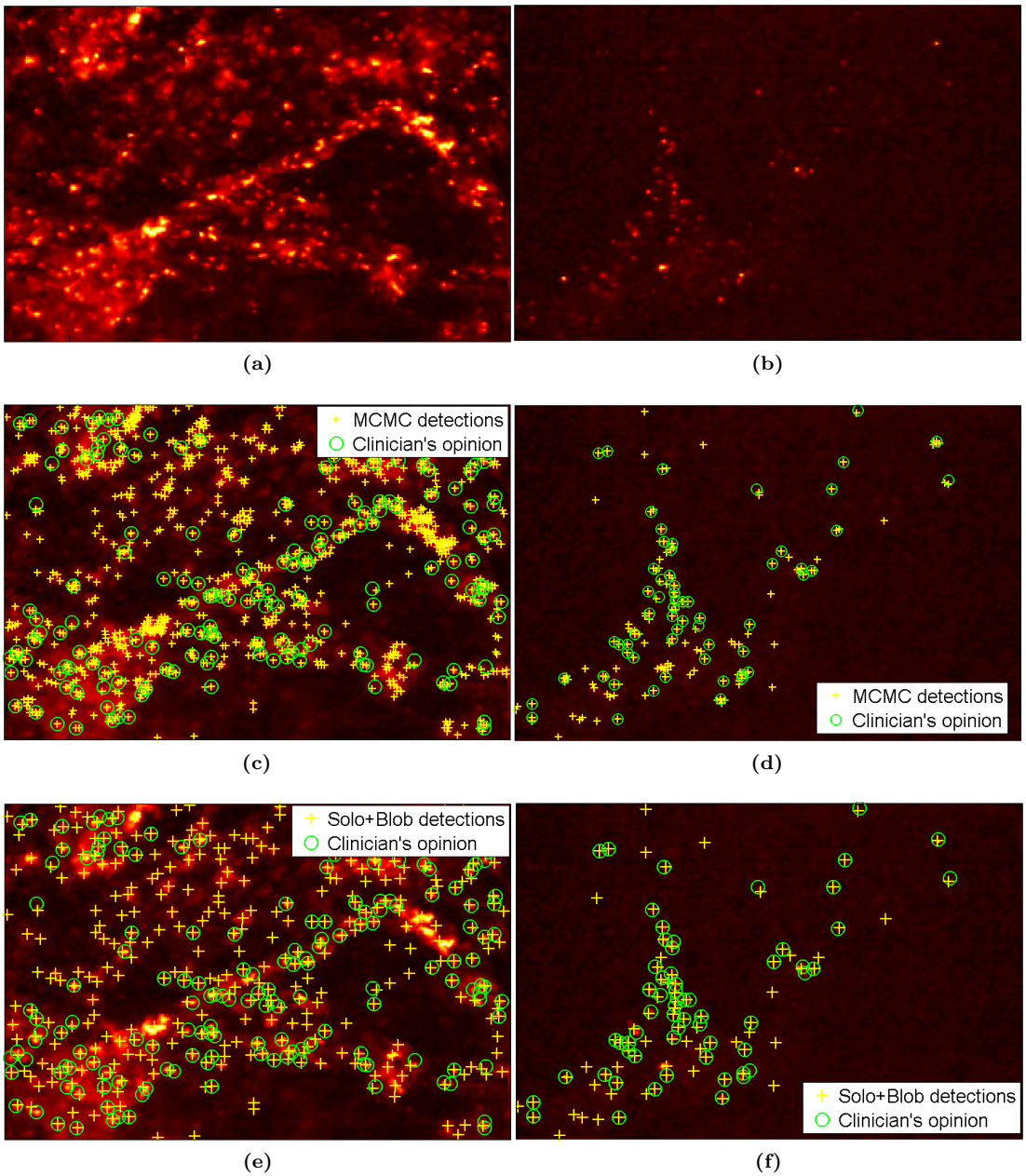


Figure 6.9 Examples of detection in two frames of Dataset I (left: video 2, right: video 5). (a) and (b) original frames, (c) and (d) all algorithm's detections along with clinician's annotations, (e) and (f) final detections after applying the post processing steps described in Section 4.7.

	Videos	Precision					Recall
		Hierarchical Bayesian model	**Sparse coding approach (Chapter 5)**	**LoG**	**DoG**	**Grey scale opening top -hat filter**	
Dataset I	1:3	35.36%	52.00%	38.01%	**54.04%**	52.03%	95.52%
	4:5	**44.97%**	19.50%	Dn/E	Dn/E	20.01%	99.00%
	6:7	**23.14%**	02.74%	0.68%	01.10%	03.31%	89.60%
	Avg.	**37.94%**	24.75%	N/A	N/A	25.14%	96.63%
Dataset II	1	**15.50%**	08.01%	03.50%	04.32%	07.53%	86.86%
	2	18.82%	**23.00%**	Dn/E	Dn/E	12.04%	98.35%
	3	17.27%	**18.01%**	Dn/E	Dn/E	09.53%	91.14%
	4	**21.63%**	19.83%	Dn/E	09.01%	11.5%	82.61%
	Avg.	**18.3%**	17.21%	N/A	N/A	10.15%	89.74%

Table 6.2 Corresponding precision measures of the patch-based approach, the LoG, the DoG and the grey scale opening top-hat filter to the Recall measures of the Bayesian spot detection algorithm. The acronyms Dn/E, N/A and Avg. refer to Doesn't exist, not applicable and average respectively.

	Hierarchical Bayesian model	**Sparse coding approach**	**LoG**	**DoG**	**Grey scale opening top-hat filter**
Computation time	5.5	$5 \times 0.4 = 2$	0.11	**0.05**	0.22

Table 6.3 Computation time (in seconds) for the two proposed methods and three existing spot-detection methods. In order to maintain a fair comparison between the five algorithms, the computation time of the sparse coding approach corresponds to the duration of five runs (used to select the best regularization parameter among the five values). Moreover, all of these methods apart from the Bayesian model requires manual thresholding of the outlier amplitude image in order to identify bacteria locations, but this can not be easily converted into precise computation time.

6.7 Conclusions

In this chapter, we have illustrated the performance of an unsupervised-based approach for bacterial detection in OEM images of distal lung tissue. The approach considered a hierarchical Bayesian model in which prior distributions were chosen for the unknown model parameters and their corresponding hyperparameters. An MCMC method based on partially collapsed Gibbs sampler was proposed to sample from the resulting posterior distribution in order to estimate the unknown model parameters. We have provided simulations on synthetic and real datasets. For synthetic data, we showed the ability of the proposed approach in denoising of fibre core intensities as well as in outlier detection using different scenarios of outlier percentages, means and variances, and noise levels. For real datasets, simulations conducted on two ovine lung datasets, including a combination of fluorescent dyes, bacterial species and bacterial loads, demonstrated that the estimated bacterial count correlates with the bacterial counts performed by a clinician and a good precision-recall values were achieved. The considered algorithm is fully automatic in the sense that it does not require the user to set any sensitive/crucial parameters as all model parameters are estimated within the MCMC algorithm. Comparisons with the unsupervised approach presented in Chapter 5 and three existing methods showed superior performance of the Bayesian approach presented in this chapter. Although the proposed algorithm in this chapter is simulation-based for which a sophisticated MCMC-based method has been considered, it provided competitive computation time to optimization-based methods, with the benefit of being fully automatic. Future work includes bacterial tracking by taking advantage of the temporal information in the datasets. Multiple concurrent imaging wavelengths can also be tested to extend to detection in multispectral datasets by using spectral unmixing algorithms.

Chapter 7

Conclusions and Suggestions for Future Work

Contents

This chapter begins with Section 7.1, which presents a broad summary of and conclusions regarding our work presented in this dissertation. Despite the advances introduced by our work, there remain some limitations and topics that could be addressed in future work. These are explored in Section 7.2.

7.1 Conclusions

We have seen in this thesis that the fibred optical endomicroscopy system has the ability to image biological tissues *in vivo* and *in situ*, and thus is a potential tool for clinicians and biologists to conduct optical biopsies in a short time. It can therefore reduce the diagnosis time of patients in intensive care units. Our goal throughout the manuscript has been to show that advanced image processing algorithms could help to overcome some hardware limitations of the fibred OEM imaging device being developed by the EPSRC-Proteus multidisciplinary project, as well as those of other similar existing devices. This research work focused on three main limitations of this

imaging technology, and provided potential solutions to each of them. First, the field of view honeycomb effect which produces pixelation artefacts; second, the fibre core cross coupling problem, and finally, the detection of fluorescently labelled bacteria which exist in a non-homogeneous background, making them difficult to detect. However, from the methodological point of view, my research work has provided a novel parametric prior blur model for the fibre core cross coupling problem; developed models for the deconvolution and restoration of irregularly subsampled images, and proposed two supervised and unsupervised-based methods for the detection of spots in non-homogeneous backgrounds. This thesis began with an overview of fibred OEM and how images are formed, then a literature review on similar problems was introduced, highlighting a lack of robust methodologies for solving these problems, and finally we introduced our main contributions in the three subsequent chapters.

The fibre core cross coupling/talk problem is a well known limitation in coherent fibre bundles. In this problem, illuminating a single fibre core will result in cross coupling at its neighbouring cores, which results in degradation of the resulting imaged structures. Moreover, the irregular packing of the fibre cores in the fibre bundle which produces cladding in between them results in pixelation artefacts, causing severe loss of information from the imaged scene. The resulting irregularly subsampled images makes it difficult to interpret the imaged scene. In Chapter 4, we provided a parametric prior blur model for this problem, in which a generalized Gaussian distribution was considered [**Eldaly et al. 2018a, Eldaly et al. 2018c**]. The parameters of this distribution depend on the coupling amount [**Perperidis et al. 2017**]. Following this, we formulated the fibre core cross coupling/talk problem through a deconvolution and restoration task for the first time in the literature. Then, we showed that the Gaussian noise model is good enough to fit this data as, from experimental analysis, it has been found that the mean of the noise is constant with respect to its variance. A complete image model was first provided. However, due to its complexity in terms of the inference step, we provided an approximation to it by representing each fibre core with a single intensity value. Due to the overlap of fibre cores, we used the maximum intensity of each fibre core due to the adequate symmetry of the distribution of the intensities of the pixels representing each core. The observation vector in this case is a concatenation of all of these intensities. The approximated image model assumed that the observations resulted from a convolutional operator representing fibre core cross coupling, acting on original fibre core intensity values, which we aim to estimate, and corrupted by observation noise, which was assumed to be zero mean i.i.d. Gaussian. Due to the ill-posed nature of the problem, some sort of regularization was required. Hence, we considered a hierarchical Bayesian model to incorporate it. In this model, suitable prior distributions were assigned to the unknown model parameters and their corresponding hyperparameters. Due to the difficulty in computing the MAP or the MMSE directly from the resulting joint posterior distribution, we introduced a sampling-based method in which we used MCMC, which was implemented using a Gibbs sampler, to infer the unknown model

parameters. Although this model is fully automatic in the sense that it can estimate all of the model parameters and hyperparameters, including regularization parameters which are difficult to tune in real world situations, it is very computationally demanding. Hence, we considered an approximation of the resulting joint posterior distribution using the variational Bayes methodology. This method not only provided a very fast estimation procedure compared to MCMC, but also provided estimations of all of the model parameters and hyperparameters. These two inference strategies were then compared with another optimization-based method, in which ADMM was considered. The rationale behind this choice was due to its wide use in the literature to solve inverse imaging problems. The three estimation algorithms showed similar performance in terms of RMSE calculation between original and estimated intensities. Several runs of the ADMM algorithm with different regularization parameters were considered to select the deconvolved samples corresponding to the lowest RMSE. Although a single run of the ADMM algorithm is faster than that of MCMC and VB, the VB algorithm was considered to be the fastest as it could provide the estimations in only a single run, with still being fully automatic. The resulting deconvolved samples were then interpolated using Gaussian processes regression in order to recover a full image. GP provided more advantage compared with other interpolation methods in that it could provide confidence intervals for the missing interpolated pixels, and hence was useful for uncertainty quantification. The three algorithms were then tested on real fibred optical endomicroscopic datasets in which they showed promising results.

In Chapters 5 and 6 we presented two unsupervised approaches for the detection of fluorescently labelled bacteria immersed in non-homogeneous backgrounds such as elastin and collagen which challenge their detection. In Chapter 5, a patch-based approach was employed in which each image is split into overlapping parches of fixed size [**Eldaly et al. 2018d**]. Each patch was then considered to result from a linear combination of background structures, corrupted by additive Gaussian noise and probably by an additional outlier term modelling anomalies (which are considered to be candidate bacteria). However, background image structures were represented by a set of atoms grouped in a dictionary learned from a separate set of frames from which patches containing bacteria were excluded. The bacteria detection task was then formulated by an optimization-based problem and the ADMM algorithm was used to estimate the unknown model parameters. This approach provided interesting results and outperformed existing methods in the literature in terms of bacteria detection performance. However, it required the user to set the regularization parameters, which was difficult due to the change of bacteria concentration in the different tested frames/datasets. Subsequently, the estimated outlier image needed to be thresholded in order to determine the final bacteria locations, which required one more parameter to tune. Moreover, due to the subjectiveness of bacteria annotation, this could result in a biased learned dictionary. Hence, Chapter 6 provided a more adaptive algorithm to deal with these limitations [**Eldaly et al. 2018b**]. This approach falls under the hierarchical Bayesian models

category. Each image was represented by a vector of fibre core intensities, each of which was then considered to result from a linear combination of the actual intensity value corresponding to background structures, corrupted by additive Gaussian noise and possibly by an additional outlier term modelling anomalies. This spot detection problem was formulated in a Bayesian framework in which suitable prior distributions were assigned to the unknown model parameters and their corresponding hyperparameters. A Markov chain Monte Carlo algorithm based on a partially collapsed Gibbs sampler was then used to sample the posterior distribution of the unknown parameters. This model is fully automatic in the sense that the user did not need to set any sensitive parameters, hence it is beneficial for datasets of different outlier concentrations. The two algorithms were tested on two real datasets, for which different combinations of SmartProbes and bacterial concentrations including control cases were considered. They were also compared with existing methods in the literature, with both showing superior performance. Regarding computational complexity, although a sampling-based method was considered for the unsupervised-based approach, which is usually very computationally expensive, it could compete with the optimization-based method considered in Chapter 5, as we could provide relatively fast implementation by taking advantage of the Markov random field property.

7.2 Suggestions for Further Work

In this manuscript, I have proposed Bayesian algorithms for restoration and bacteria detection in fibred optical endomicroscopy images. The flexibility of the proposed Bayesian frameworks allowed the model's complexity to be handled, and prior information about the model parameters to be easily included within the inference procedure. In this section, we provide future directions for those who would like to continue this work.

7.2.1 Improved computational complexity

The limitation of the proposed work in this thesis lies in the computational complexity of the developed algorithms, which limit their application to real-time processing. This indeed depends on different factors from which is the way the algorithms are implemented. For instance, conjugate gradient and diagonal approximation can be considered for efficient matrices inversion in Chapter 4 and 5. Moreover, other algorithms tailored to solving quadratic programs on the positive orthant (i.e. L-BFGS-B [112]) can be considered for the work in Chapter 4. On the other hand, low-rank sparse decompositions [35] can be used to address the bacteria detection problem solved in Chapters 5 and 6. Moreover, machine/deep learning methods can also be one of the solutions to address this limitation. However, the machine might need to be retrained for different acquisition scenarios.

7.2.2 On-line image restoration/bacteria detection

One potential way in which the deconvolution and restoration work presented in Chapter 4 could be continued would be through a consideration of temporal information in the deconvolution process. The variational Bayes approach is likely to be a good choice for this on-line deconvolution approach as it not only provides faster estimates compared to the rest of the tested algorithms, but can also estimate the model hyperparameters.

The bacteria detection work presented in Chapters 5 and 6 could be extended by taking into account the temporal information related to background image structures between successive frames. This could improve the detection performance and reduce the false detection of background image structures as bacteria.

7.2.3 Joint deconvolution and bacteria detection

In this thesis, we provided bacteria detection models for data acquired by fibred OEM systems without taking into account the cross coupling effect. Hence, in future work, one could provide models for joint deconvolution and bacteria detection, for which deconvolution would compensate for fibre core cross coupling. A potential model would be to represent each fibre core through a single intensity value as in Chapters 4 and 6, leading to an observation vector $\mathbf{y} \in \mathbb{R}^{N_1}$, where N_1 is the number of fibre cores present in the image. The observation vector in this can be represented by a convolutional operator $\mathbf{H} \in \mathbb{R}^{N_1 \times N_1}$, representing fibre core cross coupling, acting on a sum of an actual intensity value $\mathbf{x} \in \mathbb{R}^{N_1}$, representing background image structures like elastin and collagen, and a sparse outlier term modelling bacteria $\mathbf{r} \in \mathbb{R}^{N_1}$, and finally corrupted by another term modelling observation noise $\mathbf{w} \in \mathbb{R}^{N_1}$. The model then can be written as

$$\mathbf{y} = \mathbf{H}(\mathbf{x} + \mathbf{r}) + \mathbf{w}. \tag{7.1}$$

In a similar fashion to the unsupervised bacteria detection work presented in Chapter 6, a hierarchical Bayesian model can still be considered in which suitable prior distributions can be assigned to the unknown model parameters and hyperparameters, and statistical inference can be performed afterwards using either simulation or optimization-based methods.

Another potential formulation would be to consider a supervised-based approach as in Chapter 5. Each image can be split into patches, which can be collected in an observation matrix $\mathbf{Y} \in \mathbb{R}^{P \times L}$. The model would then read

$$\mathbf{Y} = \mathbf{H}(\mathbf{D}\boldsymbol{\Psi} + \mathbf{R}) + \mathbf{W}, \tag{7.2}$$

where $\mathbf{H} \in \mathbb{R}^{P \times P}$ is the fibre core cross coupling deconvolution operator, $\mathbf{D} \in \mathbb{R}^{P \times K}$ is the learned dictionary, $\boldsymbol{\Psi} \in \mathbb{R}^{K \times L}$ is the sparse coefficient matrix, $\mathbf{R} \in \mathbb{R}^{P \times L}$ has few non-zero elements that

equal the deviation of each outlier from the sparse representations model, and $\mathbf{W} \in \mathbb{R}^{P \times K}$ is a low-energy noise component.

However, this would require learning the dictionary $\mathbf{D}$ from deconvolved and outlier free images, for which any of the algorithms presented in Chapter 4 can be considered.

7.2.4 Joint spectral unmixing and anomaly detection in multispectral datasets

The imaging device being developed within the EPSRC-Proteus project is currently being extended for multispectral excitation and collection, which can in turn provide multispectral images of the same scene. Spectral unmixing refers to any process that separates the pixel spectra from a multispectral/hyperspectral image into a collection of constituent spectra, or spectral signatures, called endmembers, and a set of fractional abundances, one set per pixel. The endmembers are generally assumed to represent the pure materials present in the image, and the set of abundances at each pixel to represent the percentage of each endmember that is present in the pixel. This would be useful to identify the materials constituting the tissues. Moreover, this can also be extended to joint material identification and outlier detection. A potential observation model would be

$$\mathbf{y} = \mathbf{B}\mathbf{a} + \mathbf{r} + \mathbf{w}, \tag{7.3}$$

where $\mathbf{y}_n = [y_{1,n}, \cdots, y_{L,n}]$, where $n \in \{1, \cdots, N\}$, where N is the total number of observed pixels/spectra in the image (or N_1 if each fibre core is represented by a single intensity value), and L is the number of spectral bands. Each of these spectra is assumed to result from a linear combination of $\mathbf{B}$ endmembers, weighted by their proportions $\mathbf{a}$, corrupted by possible additive outliers $\mathbf{r}$ and Gaussian noise $\mathbf{w}$.

7.2.5 Other inference strategies

The Bayesian model considered in Chapter 6 for the detection of fluorescently labelled bacteria can take advantage of the sparsity of the outlier vector and consider a reversible jump Markov chain Monte Carlo method [79] to infer the unknown parameters. This algorithm is an extension to standard MCMC methodology that allows simulation of the posterior distribution on spaces of varying dimensions. The change in dimensionality comes from the fact that the observation model for the detection of outliers can be written as

$$y_n = \begin{cases} y_n = x_n + t_n + w_n & \text{when } z_n = 1, \\ y_n = x_n + w_n & \text{when } z_n = 0. \end{cases} \tag{7.4}$$

The reversible jump MCMC algorithm "jumps" between these two models, which are of different dimensionality, until convergence.

Bibliography

[1] Muriel Abbaci, Ingrid Breuskin, Odile Casiraghi, Frederic De Leeuw, Malek Ferchiou, Stephane Temam, and Corinne Laplace-Builhé. Confocal laser endomicroscopy for non-invasive head and neck cancer imaging: a comprehensive review. *Oral oncology*, 50(8):711–716, June 2014.

[2] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. on Image Process.*, 19(9):2345–2356, Sept 2010.

[3] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. on Image Process.*, 20(3):681–695, March 2011.

[4] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Non-cyclic deconvolution using an augmented lagrangian method. In *IEEE International Conference on Computer as a Tool (EUROCON)*, pages 1–4, April 2011.

[5] Michal Aharon, Michael Elad, Alfred Bruckstein, et al. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Process.*, 54(11):4311, Nov 2006.

[6] Ahsan R Akram, Nicolaos Avlonitis, Thomas Craven, Marc Vendrell, Neil McDonald, Emma Scholefield, Andrew Fisher, Paul Corris, Christopher Haslett, Mark Bradley, and Kevin Dhaliwal. Structural modifications of the antimicrobial peptide ubiquicidin for pulmonary imaging of bacteria in the alveolar space. *The Lancet*, 387:S17, Feb 2016. Spring Meeting for Clinician Scientists in Training 2016.

[7] Ahsan R Akram, Nicolaos Avlonitis, Annamaria Lilienkampf, Ana M Perez-Lopez, Neil McDonald, Sunay V Chankeshwara, Emma Scholefield, Christopher Haslett, Mark Bradley, and Kevin Dhaliwal. A labelled-ubiquicidin antimicrobial peptide for immediate in situ optical detection of live bacteria in human alveolar lung tissue. *Chemical Science*, 6(12):6971–6979, Jun 2015.

[8] Mariana SC Almeida and Mario Figueiredo. Deconvolving images with unknown boundaries using the alternating direction method of multipliers. *IEEE Trans. on Image Process.*, 22(8):3074–3086, Aug 2013.

[9] Yoann Altmann, Stephen McLaughlin, and Alfred Hero. Robust linear spectral unmixing using anomaly detection. *IEEE Trans. Comput. Imag.*, 1(2):74–85, June 2015.

[10] Yoann Altmann, Steve McLaughlin, and Alfred Hero. Robust linear spectral unmixing using outlier detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 74–85, June 2015.

[11] Carlos Arteta, Victor Lempitsky, J Alison Noble, and Andrew Zisserman. Learning to detect cells using non-overlapping extremal regions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 348–356. Springer, 2012.

[12] Carlos Arteta, Victor Lempitsky, J Alison Noble, and Andrew Zisserman. Interactive object counting. pages 504–518, September 2014.

[13] Carlos Arteta, Victor Lempitsky, and Andrew Zisserman. Counting in the wild. In *European Conference on Computer Vision (ECCV)*, pages 483–498, Amsterdam, The Netherlands, Oct 2016. Springer.

[14] Tashfeen Aslam, Amy Miele, Sunay V Chankeshwara, Alicia Megia-Fernandez, Chesney Michels, Ahsan R Akram, Neil McDonald, Nik Hirani, Chris Haslett, Mark Bradley, et al. Optical molecular imaging of lysyl oxidase activity–detection of active fibrogenesis in human lung tissue. *Chemical Science*, 6(8):4946–4953, 2015.

[15] Nicolaos Avlonitis, Manuelle Debunne, Tashfeen Aslam, Neil McDonald, Chris Haslett, Kevin Dhaliwal, and Mark Bradley. Highly specific, multi-branched fluorescent reporters for analysis of human neutrophil elastase. *Organic & biomolecular chemistry*, 11(26):4414–4418, May 2013.

[16] Nicholas Ayache, Tom Vercauteren, Gregoire Malandain, Fabien Oberrietter, Nicolas Savoire, and Aymeric Perchant. Processing and mosaicing of fibered confocal images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI): Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB)*, pages 1–5, Copenhagen, Denmark, 2006. Springer. Invited.

[17] Gerald Ayers and J Christopher Dainty. Iterative blind deconvolution method and its applications. *Optics letters*, 13(7):547–549, July 1988.

[18] Derin Babacan, Rafael Molina, and Aggelos Katsaggelos. Variational bayesian super resolution. *IEEE Trans. on Image Process.*, 20(4):984–999, April 2011.

[19] Derin Babacan, Rafael Molina, and Aggelos K Katsaggelos. Parameter estimation in tv image restoration using variational distribution approximation. *IEEE Trans. on Image Process.*, 17(3):326–339, March 2008.

[20] Derin Babacan, Rafael Molina, and Aggelos K Katsaggelos. Total variation super resolution using a variational approach. In *IEEE International Conference on Image Processing (ICIP)*, pages 641–644, Oct 2008.

[21] Derin Babacan, Rafael Molina, and Aggelos K Katsaggelos. Variational bayesian blind deconvolution using a total variation prior. *IEEE Trans. on Image Process.*, 18(1):12–26, Jan 2009.

[22] Vickie S Baselski and Richard G Wunderink. Bronchoscopic diagnosis of pneumonia. *Clinical microbiology reviews*, 7(4):533–558, Oct. 1994.

[23] Matthew James Beal. *Variational algorithms for approximate Bayesian inference*. University of London United Kingdom, 2003.

[24] Murat Belge, Misha E Kilmer, and Eric L Miller. Wavelet domain image restoration with adaptive edge-preserving regularization. *IEEE Trans. on Image Process.*, 9(4):597–608, April 2000.

[25] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[26] Dhruba J Bharali, Ilona Klejbor, Ewa K Stachowiak, Purnendu Dutta, Indrajit Roy, Navjot Kaur, Earl J Bergey, Paras N Prasad, and Michal K Stachowiak. Organically modified silica nanoparticles: a nonviral vector for in vivo gene delivery and expression in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32):11539–11544, Aug 2005.

[27] José M Bioucas-Dias, Mário AT Figueiredo, and Joao P Oliveira. Adaptive total variation image deconvolution: A majorization-minimization approach. In *IEEE European Signal Processing Conference (EUSIPCO)*, pages 1–4, Florence, Italy, Sept. 2006.

[28] Tom E Bishop, Derin Babacan, Bruno Amizic, Aggelos K Katsaggelos, Tony Chan, and Rafael Molina. Blind image deconvolution: problem formulation and existing approaches. *Blind image deconvolution: Theory and Applications*, pages 1–41, 2007.

[29] Tom E Bishop and James R Hopgood. Blind image restoration using a block-stationary signal model. In *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, volume 2, May 2006.

[30] Tom E Bishop, Rafael Molina, and James R Hopgood. Blind restoration of blurred photographs via ar modelling and mcmc. In *IEEE International Conference on Image Processing (ICIP)*, pages 669–672, San Diego, CA, USA, Oct 2008.

[31] Charles Bouman and Ken Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Trans. on Image Process.*, 2(3):296–310, July 1993.

[32] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[33] David S Bright and Eric B Steel. Two-dimensional top hat filter for extracting spots and spheres from digital images. *Journal of Microscopy*, 146(2):191–200, May 1987.

[34] Jacobus Burggraaf, Ingrid MC Kamerling, Paul B Gordon, Lenneke Schrier, Marieke L De Kam, Andrea J Kales, Ragnar Bendiksen, Bård Indrevoll, Roger M Bjerke, Siver A Moestue, et al. Detection of colorectal polyps in humans using an intravenously administered fluorescent peptide targeted against c-met. *Nature medicine*, 21(8):955–961, Aug. 2015.

[35] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[36] Mark B Cannell, Angus McMorland, and Christian Soeller. Image enhancement by deconvolution. *Handbook of biological confocal microscopy*, pages 488–500, Oct. 2006.

[37] Michael Cannon. Blind deconvolution of spatially invariant image blurs with phase. *IEEE Trans. on Acoustics, Speech, and Signal Process.*, 24(1):58–63, Feb 1976.

[38] Tony Chan and Jianhong Shen. *Image Processing and Analysis-Variational, PDE, Wavelet, and Stochastic Methods Society for Industrial and Applied Mathematics.* 2005.

[39] Tony F Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE Trans. on Image Process.*, 7(3):370–375, March 1998.

[40] Timothy C Chang, Jen-Jane Liu, and Joseph C Liao. Probe-based confocal laser endomicroscopy of the urinary tract: the technique. *Journal of visualized experiments: JoVE*, (71):1–1, 2013.

[41] Giannis Chantas, Nikolaos P Galatsanos, Rafael Molina, and Aggelos K Katsaggelos. Variational bayesian image restoration with a product of spatially weighted total variation image priors. *IEEE Trans. on Image Process.*, 19(2):351–362, Feb 2010.

[42] Jean Chastre and Jean-Yves Fagon. Ventilator-associated pneumonia. *American journal of respiratory and critical care medicine*, 165(7):867–903, 2002.

[43] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, Feb. 2001.

[44] Stephanie P Chen and Joseph C Liao. Confocal laser endomicroscopy of bladder and upper tract urothelial carcinoma: a new era of optical diagnosis? *Current urology reports*, 15(9):437, July 2014.

[45] Xianpei Chen, Kristen Lantz Reichenbach, and Chris Xu. Experimental and theoretical analysis of core-to-core coupling on fiber bundle imaging. *Optics express*, 16(26):21598–21607, Dec. 2008.

[46] Brent A Chipman and Brian D Jeffs. Blind multiframe point source image restoration using map estimation. In *Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers*, volume 2, pages 1267–1271. IEEE, Oct 1999.

[47] Emilie Chouzenoux, Anna Jezierska, Jean-Christophe Pesquet, and Hugues Talbot. A convex approach for image restoration with exact poisson–gaussian likelihood. *SIAM Journal on Imaging Sciences*, 8(4):2662–2682, 2015.

[48] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 411–418. Springer, 2013.

[49] Patrick L Combettes and Valérie R Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, Nov 2005.

[50] Barham K Abu Dayyeh, Nirav Thosani, Vani Konda, Michael B Wallace, Douglas K Rex, Shailendra S Chauhan, Joo Ha Hwang, Sri Komanduri, Michael Manfredi, John T Maple, et al. Asge technology committee systematic review and meta-analysis assessing the asge pivi thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointestinal endoscopy*, 81(3):502–e1, Dec 2015.

[51] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 39(1):1–38, 1977.

[52] Haluk Derin and Howard Elliott. Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, (1):39–55, Jan 1987.

[53] Xinghao Ding, Lihan He, and Lawrence Carin. Bayesian robust principal component analysis. *IEEE Trans. on Image Process.*, 20(12):3419–3430, Dec 2011.

[54] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Trans. on Image Process.*, 20(7):1838–1857, July 2011.

[55] James East, Jasper L Vleugels, Philip Roelandt, Pradeep Bhandari, Raf Bisschops, Evelien Dekker, Cesare Hassan, Gareth Horgan, Ralf Kiesslich, Gaius Longcroft-Wheaton, et al. Advanced endoscopic imaging: European society of gastrointestinal endoscopy (esge) technology review. Nov. 2016.

[56] Jonathan Eckstein and Dimitri P Bertsekas. On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, Apr 1992.

[57] Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[58] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. on Image Process.*, 15(12):3736–3745, Dec 2006.

[59] Ahmed Karam Eldaly, Y. Altmann, A. Perperidis, N. Krstajic, T. R. Choudhary, K. Dhaliwal, and S. McLaughlin. Deconvolution and restoration of optical endomicroscopy images. *IEEE Trans. Comput. Imag.*, 4(2):194–205, March 2018.

[60] Ahmed Karam Eldaly, Yoann Altmann, Antonios Perperidis, and Stephen McLaughlin. Deconvolution of irregularly subsampled images. In *IEEE Statistical Signal Processing Workshop (SSP)*, pages 303–307, Freiburg, Germany, June 2018.

[61] Kjersti Engan, Sven Ole Aase, and John Håkon Husøy. Multi-frame compression: Theory and design. *Signal Processing*, 80(10):2121–2140, Oct. 2000.

[62] Jalal M Fadili and Jean-Luc Starck. Sparse representation-based image deconvolution by iterative thresholding. In *Astronomical Data Analysis (ADA)*, volume 1, pages 1–10, Marseille, France, Jan. 2006.

[63] Houzhang Fang, Chunan Luo, Gang Zhou, and Xiaoping Wang. Hyperspectral image deconvolution with a spectral-spatial total variation regularization. *Canadian Journal of Remote Sensing*, 43(4):384–395, Aug 2017.

[64] Mário AT Figueiredo and José M Bioucas-Dias. Restoration of poissonian images using alternating direction optimization. *IEEE Trans. on Image Process.*, 19(12):3133–3145, Dec 2010.

[65] Mario AT Figueiredo, Jose M Bioucas-Dias, and Manya V Afonso. Fast frame-based image deconvolution using variable splitting and constrained optimization. In *IEEE Workshop on Statistical Signal Processing (SSP)*, pages 109–112, Cardiff, United Kingdom, July 2009.

[66] Mário AT Figueiredo and Robert D Nowak. Wavelet-based image estimation: an empirical bayes approach using jeffrey's noninformative prior. *IEEE Trans. on Image Process.*, 10(9):1322–1331, Sept 2001.

[67] Benjamin A Flusberg, Eric D Cocker, Wibool Piyawattanametha, Juergen C Jung, Eunice LM Cheung, and Mark J Schnitzer. Fiber-optic fluorescence imaging. *Nature methods*, 2(12):941–950, Nov. 2005.

[68] Benjamin A Flusberg, Axel Nimmerjahn, Eric D Cocker, Eran A Mukamel, Robert PJ Barretto, Tony H Ko, Laurie D Burns, Juergen C Jung, and Mark J Schnitzer. High-speed, miniaturized fluorescence microscopy in freely moving mice. *Nature methods*, 5(11):935–938, Oct 2008.

[69] Coralie Fouquet, Jean-François Gilles, Nicolas Heck, Marc Santos, Richard Schwartzmann, Vidjeacoumary Cannaya, Marie-Pierre Morel, Robert Stephen Davidson, Alain Trembleau, and Susanne Bolte. Improving axial resolution in confocal microscopy with new high refractive index mounting media. *PLoS One*, 10(3):1–17, March 2015.

[70] Florian Fuchs, Sabine Zirlik, Kai Hildner, Juergen Schubert, Michael Vieth, and Markus Neurath. Confocal laser endomicroscopy for diagnosing lung cancer in vivo. *European Respiratory Journal*, 41(6):1401–1408, Sept 2013.

[71] Alessandro Fugazza, Federica Gaiani, Maria Clotilde Carra, Francesco Brunetti, Michaël Lévy, Iradj Sobhani, Daniel Azoulay, Fausto Catena, Gian Luigi de'Angelis, and Nicola de'Angelis. Confocal laser endomicroscopy in gastrointestinal and pancreatobiliary diseases: a systematic review and meta-analysis. *BioMed research international*, pages 1–31, 2016.

[72] Nikolas P Galatsanos, Vladimir Z Mesarovic, Rafael Molina, and Aggelos K Katsaggelos. Hierarchical bayesian image restoration from partially known blurs. *IEEE Trans. on Image Process.*, 9(10):1784–1797, Oct 2000.

[73] Nikolas P Galatsanos, Vladimir Z Mesarovic, Rafael Molina, Aggelos K Katsaggelos, and Javier Mateos. Hyperparameter estimation in image restoration problems with partially-known blurs. *Optical Engineering*, 41(8):1845–1855, Aug. 2002.

[74] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL, 2014.

[75] Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[76] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *Readings in Computer Vision*, volume PAMI-6, pages 721–741. Elsevier, Nov 1987.

[77] Daniel Gerlich and Jan Ellenberg. 4d imaging to assay complex dynamics in live specimens. *Nature cell biology*, pages 14–19, Sept 2003.

[78] Dong Gong, Mingkui Tan, Yanning Zhang, Anton Van den Hengel, and Qinfeng Shi. Blind image deconvolution by automatic gradient activation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1827–1836, Las Vegas, United States, June 2016.

[79] Peter J Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, June 1995.

[80] Per Christian Hansen, James G Nagy, and Dianne P O'leary. *Deblurring images: matrices, spectra, and filtering*, volume 3. Siam, 2006.

[81] Feng He, Bangshu Xiong, Chengli Sun, and Xiaobin Xia. A laplacian of gaussian-based approach for spot detection in two-dimensional gel electrophoresis images. In *International Conference on Computer and Computing Technologies in Agriculture*, pages 8–15, Beijing, China, Sept. 2010. Springer.

[82] Xiangqian Hong, Vivek K Nagarajan, Dale H Mugler, and Bing Yu. Smartphone microendoscopy for high resolution fluorescence imaging. *Journal of Innovative Optical Health Sciences*, 9(05):1–4, March 2016.

[83] Pei-Lin Hsiung, Jonathan Hardy, Shai Friedland, Roy Soetikno, Christine B Du, Amy P Wu, Peyman Sahbaie, James M Crawford, Anson W Lowe, Christopher H Contag, et al. Detection of colonic dysplasia in vivo using a targeted heptapeptide and confocal microendoscopy. *Nature medicine*, 14(4):454–458, March 2008.

[84] Joey M Jabbour, Meagan A Saldua, Joel N Bixler, and Kristen C Maitland. Confocal endomicroscopy: instrumentation and medical applications. *Annals of biomedical engineering*, 40(2):378–397, Oct. 2012.

[85] Edwin T Jaynes. Prior probabilities. *IEEE Trans. on Systems Science and Cybernetics*, 4(3):227–241, 1968.

[86] Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):453–461, Sept. 1946.

[87] Shan Jiang, Xiaobo Zhou, Tom Kirchhausen, and Stephen TC Wong. Detection of molecular particles in live cells via machine learning. *Cytometry Part A*, 71(8):563–575, Aug 2007.

[88] Radovan Jirik and Torfinn Taxt. Two-dimensional blind bayesian deconvolution of medical ultrasound images. *IEEE Trans. on Ultrasonics, ferroelectrics, and frequency control*, 55(10), Oct 2008.

[89] Paul Johnston, Danny F McAuley, and Cecilia M O'Kane. Novel pulmonary biomarkers in the diagnosis of vap. *Thorax*, 65(3):190–192, March 2010.

[90] Kyung Sub Joo, Tamal Bose, and Guo Fang Xu. Image restoration using a conjugate gradient-based adaptive filtering algorithm. *Circuits, Systems and Signal Processing*, 16(2):197–206, 1997.

[91] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[92] Kunal Karia and Michel Kahaleh. A review of probe-based confocal laser endomicroscopy for pancreaticobiliary disease. *Clinical endoscopy*, 49(5):462–466, 2016.

[93] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, March 1995.

[94] Aggelos K Katsaggelos and Kuen-Tsair Lay. Maximum likelihood blur identification and image restoration using the em algorithm. *IEEE Trans. on Signal Process.*, 39(3):729–733, March 1991.

[95] Supang Khondee and Thomas D Wang. Progress in molecular imaging in endoscopy and endomicroscopy for cancer imaging. *Journal of healthcare engineering*, 4(1):1–22, Dec 2013.

[96] Yoshitaka Kimori, Norio Baba, and Nobuhiro Morone. Extended morphological processing: a practical method for automatic spot detection of biological markers from microscopic images. *BMC bioinformatics*, 11(1):1–13, July 2010.

[97] Hui Kong, Hatice Cinar Akakin, and Sanjay E Sarma. A generalized laplacian of gaussian filter for blob detection and its applications. *IEEE transactions on cybernetics*, 43(6):1719–1733, 2013.

[98] Nikola Krstajić, Ahsan R Akram, Tushar R Choudhary, Neil McDonald, Michael G Tanner, Ettore Pedretti, Paul A Dalgarno, Emma Scholefield, John M Girkin, Anne Moore, et al. Two-color widefield fluorescence microendoscopy enables multiplexed molecular imaging in the alveolar space of human lung tissue. *Journal of Biomedical Optics*, 21(4):046009–046009, 2016.

[99] Solomon Kullback. *Information theory and statistics*. Courier Corporation, 1997.

[100] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[101] Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution. *IEEE Signal Process. Magazine*, 13(3):43–64, May 1996.

[102] Elisabeth Laemmel, Magalie Genet, Georges Le Goualher, Aymeric Perchant, Jean-François Le Gargasson, and Eric Vicaut. Fibered confocal fluorescence microscopy (cell-vizio™) facilitates extended imaging in the field of microcirculation. *Journal of vascular research*, 41(5):400–411, Sept. 2004.

[103] Reginald L Lagendijk, Jan Biemond, and Dick E Boekee. Identification and restoration of noisy blurred images using the expectation-maximization algorithm. *IEEE Trans. on Acoustics, Speech, and Signal Process.*, 38(7):1180–1191, July 1990.

[104] Kuen-Tsair Lay and Aggelos K Katsaggelos. Image identification and restoration based on the expectation-maximization algorithm. *Optical Engineering*, 29(5):436–446, May 1990.

[105] Georges Le Goualher, Aymeric Perchant, Magalie Genet, Charlotte Cavé, Bertrand Viellerobe, Fredéric Berier, Benjamin Abrat, and Nicholas Ayache. Towards optical biopsies with an integrated fibered confocal fluorescence microscope. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 761–768, Saint-Malo, Brittany, France, Sept 2004. Springer.

[106] Hsien-Che Lee. Review of image-blur models in a photographic system using the principles of optics. *Optical Engineering*, 29(5):405–422, 1990.

[107] John Lee and Christopher J Rozell. Fast admm solver for reweighted total variation image deconvolution and inpainting. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP),*, pages 1145–1149, Montreal, Canada, Nov 2017.

[108] Victor Lempitsky and Andrew Zisserman. Learning to count objects in images. In *Advances in neural information processing systems*, pages 1324–1332, Dec 2010.

[109] Boris Lenseigne, Priscille Brodin, Hee Kyoung Jeon, Thierry Christophe, and Auguste Genovesio. Support vector machines for automatic detection of tuberculosis bacteria in confocal

microscopy images. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 85–88, Arlington, VA, USA, April 2007.

[110] Aristidis C Likas and Nikolas P Galatsanos. A variational approach for bayesian blind image deconvolution. *IEEE Trans. on Signal Process.*, 52(8):2222–2233, Aug 2004.

[111] Tony Lindeberg. Feature detection with automatic scale selection. *International journal of computer vision*, 30(2):79–116, July 1998.

[112] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[113] Jun S Liu. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966, Sept 1994.

[114] Xuan Liu, Yong Huang, and Jin Ung Kang. Dark-field illuminated reflectance fiber bundle endoscopic microscope. *Journal of biomedical optics*, 16(4):1–7, April 2011.

[115] R Lokhande, KV Arya, and P Gupta. Identification of parameters and restoration of motion blurred images. In *Proceedings of the 2006 ACM symposium on Applied computing*, volume 1, pages 301–305. ACM, April 2006.

[116] Martin Luessi, Derin Babacan, Rafael Molina, James R Booth, and Aggelos K Katsaggelos. Variational bayesian causal connectivity analysis for fmri. *Frontiers in neuroinformatics*, 8(45):1–16, May 2014.

[117] Liyan Ma, Lionel Moisan, Jian Yu, and Tieyong Zeng. A dictionary learning approach for poisson image deblurring. *IEEE Trans. on Medical Imag.*, 32(7):1277–1289, July 2013.

[118] David JC MacKay. Probable networks and plausible predictions—a review of practical bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3):469–505, Feb 1995.

[119] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[120] Stéphane Mallat and Zhifeng Zhang. Matching pursuit with time-frequency dictionaries. Technical report, Courant Institute of Mathematical Sciences New York United States, 1993.

[121] Ondřej Mandula, Ivana Šumanovac Šestak, Rainer Heintzmann, and Christopher KI Williams. Localisation microscopy with quantum dots using non-negative matrix factorisation. *Optics express*, 22(20):24594–24605, Sept. 2014.

[122] Kanti Mardia. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, 24(2):265–284, Feb. 1988.

[123] Vincent Mazet, David Brie, and Jérôme Idier. Simulation of postive normal variables using several proposal distributions. In *IEEE Statistical Signal Processing Workshop (SSP)*, pages 37–42, Bordeaux, France, July 2005.

[124] Paul McCool, Yoann Altmann, Antonios Perperidis, and Stephen McLaughlin. Robust Markov random field outlier detection and removal in subsampled images. In *IEEE Statistical Signal Processing Workshop (SSP)*, pages 1–5, Palma de Mallorca, Spain, June 2016.

[125] Erik Meijering, Oleh Dzyubachyk, and Ihor Smal. Methods for cell and particle tracking. In *Methods in enzymology*, volume 504, pages 183–200. Elsevier, Jan 2012.

[126] Oleg V Michailovich and Dan Adam. A novel approach to the 2-d blind deconvolution problem in medical ultrasound. *IEEE Trans. on Medical Imag.*, 24(1):86–104, Jan 2005.

[127] James W Miskin. Ensemble learning for independent component analysis. In *In Advances in Independent Component Analysis*. Citeseer, 2000.

[128] Rafael Molina. On the hierarchical bayesian approach to image restoration: applications to astronomical images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(11):1122–1128, Nov 1994.

[129] Rafael Molina, Antonio López, José Manuel Martín, and Aggelos K Katsaggelos. Variational posterior distribution approximation in bayesian emission tomography reconstruction using a gamma mixture prior. In *VISAPP*, pages 165–176, Barcelona, Spain, March 2007.

[130] Rafael Molina, Javier Mateos, and Aggelos K Katsaggelos. Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Trans. on Image Process.*, 15(12):3715–3727, Dec 2006.

[131] Rafael Molina and BD Ripley. Using spatial models as priors in astronomical image analysis. *Journal of Applied Statistics*, 20(5-6):281–298, April 1993.

[132] Michael A Mooney, Aqib H Zehri, Joseph F Georges, and Peter Nakaji. Laser scanning confocal endomicroscopy in the neurosurgical operating room: a review and discussion of future applications. *Neurosurgical focus*, 36(2):E9, 2014.

[133] Rahul Mourya, Loic Denis, Jean-Marie Becker, and Eric Thiébaut. Augmented lagrangian without alternating directions: Practical algorithms for inverse problems in imaging. In *IEEE International Conference on Image Processing (ICIP)*, pages 1205–1209, Quebec, Canada, Sept. 2015.

[134] Ryo Nakagaki and Aggelos K Katsaggelos. A vq-based blind image restoration algorithm. *IEEE Trans. on image Process.*, 12(9):1044–1053, Sept 2003.

[135] Radford M Neal. Probabilistic inference using markov chain monte carlo methods. pages 1–144, 1993.

[136] Gregory E. Newstadt, Alfred O. Hero, and J. Simmons. Robust spectral unmixing for anomaly detection. In *IEEE Statistical Signal Processing Workshop (SSP)*, pages 109–112, Gold Coast, Australia, June 2014.

[137] Richard Newton, Samuel Kemp, Guang-Zhong Yang, Daniel Elson, Ara Darzi, and Pallav Shah. Imaging parenchymal lung diseases with confocal endomicroscopy. *Respiratory medicine*, 106(1):127–137, Jan 2012.

[138] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[139] Gyungseok Oh, Euiheon Chung, and Seok H Yun. Optical fibers for high-resolution in vivo microendoscopic fluorescence imaging. *Optical Fiber Technology*, 19(6):760–771, Aug. 2013.

[140] Joao P Oliveira, Mario AT Figueiredo, and Jose M Bioucas-Dias. Parametric blur estimation for blind restoration of natural images: linear motion and out-of-focus. *IEEE Trans. on Image Process.*, 23(1):466–477, Jan 2014.

[141] François Orieux, Jean-François Giovannelli, and Thomas Rodet. Bayesian estimation of regularization and point spread function parameters for wiener–hunt deconvolution. *JOSA A*, 27(7):1593–1607, June 2010.

[142] Ari Pakman and Liam Paninski. Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542, Aug 2014.

[143] Ying Pan, Jens-Peter Volkmer, Kathleen E Mach, Robert V Rouse, Jen-Jane Liu, Debashis Sahoo, Timothy C Chang, Thomas J Metzner, Lei Kang, Matt Van De Rijn, et al. Endoscopic molecular imaging of human bladder cancer using a cd47 antibody. *Science translational medicine*, 6(260):1–9, Oct 2014.

[144] Kannan Panchapakesan, David G Sheppard, Michael W Marcellin, and Bobby R Hunt. Blur identification from vector quantizer encoder distortion. *IEEE Trans. on Image Procss.*, 10(3):465–470, Oct 2001.

[145] Giorgio Parisi and Ramamurti Shankar. Statistical field theory, 1988.

[146] Antonios Perperidis, Helen E Parker, Ahmed Karam-Eldaly, Yoann Altmann, Kevin Dhaliwal, Robert R Thomson, Michael G Tanner, and Stephen McLaughlin. Characterization and modelling of inter-core coupling in coherent fiber bundles. *Optics Express*, 25(10):11932–11953, May 2017.

[147] Mark Pierce, Dihua Yu, and Rebecca Richards-Kortum. High-resolution fiber-optic microendoscopy for in situ cellular imaging. *Journal of visualized experiments (JoVE)*, (47), 2011.

[148] Jing Qin, Xiyu Yi, Shimon Weiss, and Stanley Osher. Shearlet-tgv based fluorescence microscopy image deconvolution. *CAM Report. University of California, Los Angeles (UCLA)*, pages 14–32, 2014.

[149] Howard Raiffa. *Applied statistical decision theory*. Div. of Research, Graduate School of Business Administration, Harvard Univ., 1974.

[150] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, 2006.

[151] Ditlev Nytoft Rasmussen, John Gásdal Karstensen, Lene Buhl Riis, Jørn Brynskov, and Peter Vilmann. Confocal laser endomicroscopy in inflammatory bowel disease–a systematic review. *Journal of Crohn's and Colitis*, 9(12):1152–1159, July 2015.

[152] Kristen Lantz Reichenbach and Chris Xu. Numerical analysis of light propagation in image fibers or coherent fiber bundles. *Optics express*, 15(5):2151–2165, March 2007.

[153] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.

[154] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[155] Joseph JK O Ruanaidh and William J Fitzgerald. *Numerical Bayesian methods applied to signal processing*. Springer Science & Business Media, 2012.

[156] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, June 2010.

[157] Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

[158] Pablo Ruiz, Xu Zhou, Javier Mateos, Rafael Molina, and Aggelos K Katsaggelos. Variational bayesian blind image deconvolution: A review. *Digital Signal Processing*, 47:116–127, May 2015.

[159] Francesca Salvatori, Saverio Siciliano, Francesco Maione, Dario Esposito, Stefania Masone, Marcello Persico, and Giovanni D De Palma. Confocal laser endomicroscopy in the study of colonic mucosa in ibd patients: a review. *Gastroenterology research and practice*, 2012:1–7, 2012.

[160] Pinaki Sarder and Arye Nehorai. Deconvolution methods for 3-d fluorescence microscopy images. *IEEE Signal Process. Magazine*, 23(3):32–45, May 2006.

[161] Nicolas Savoire, Barbara André, and Tom Vercauteren. Online blind calibration of non-uniform photodetectors: Application to endomicroscopy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 639–646, Nice, France, 2012. Springer.

[162] Richard R Schultz and Robert L Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Trans. on Image Process.*, 5(6):996–1011, June 1996.

[163] Juan G Serra, Javier Mateos, Rafael Molina, and Aggelos K Katsaggelos. Parameter estimation in spike and slab variational inference for blind image deconvolution. In *IEEE European Signal Processing Conference (EUSIPCO)*, pages 1495–1499, Greece Kos island, Aug 2017.

[164] Sohan Seth, Ahsan R Akram, Kevin Dhaliwal, and Christopher KI Williams. Estimating bacterial and cellular load in fcfm imaging. *Journal of Imaging*, 4(1):1–11, Jan 2018.

[165] Dongsuk Shin, Mark C Pierce, Ann M Gillenwater, Michelle D Williams, and Rebecca R Richards-Kortum. A fiber-optic fluorescence microscope using a consumer-grade digital camera for in vivo cellular imaging. *PLoS One*, 5(6):e11218, June 2010.

[166] Robin Sibson. A brief description of natural neighbor interpolation. *Interpreting multivariate data*, pages 21–36, 1981.

[167] SilverFast. Silverfast resolution target (usaf 1951) by lasersoft imaging. *available online http://www.silverfast.com/show/resolution-target/en.html*, 2012.

[168] Miguel Simões, Luis B Almeida, Jose Bioucas-Dias, and Jocelyn Chanussot. A framework for fast image deconvolution with incomplete observations. *IEEE Trans. on Image Process.*, 25(11):5266–5280, Nov 2016.

[169] Ihor Smal, Marco Loog, Wiro Niessen, and Erik Meijering. Quantitative comparison of spot detection methods in fluorescence microscopy. *IEEE Trans. on Medical Imag.*, 29(2):282–301, Feb 2010.

[170] Ihor Smal, Erik Meijering, Katharina Draegestein, Niels Galjart, Ilya Grigoriev, Anna Akhmanova, ME Van Royen, Adriaan B Houtsmuller, and Wiro Niessen. Multiple object

tracking in molecular bioimaging by rao-blackwellized marginal particle filtering. *Medical Image Analysis*, 12(6):764–777, Dec 2008.

[171] Ihor Smal, Wiro Niessen, and Erik Meijering. A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 264–267, Paris, France, May 2008.

[172] Václav Šmídl and Anthony Quinn. *The variational Bayes method in signal processing*. Springer Science & Business Media, 2006.

[173] Ioana Smith, Pamela E Kline, Monica Gaidhane, and Michel Kahaleh. A review on the use of confocal laser endomicroscopy in the bile duct. *Gastroenterology research and practice*, 2012:1–5, April 2012.

[174] Allan W Snyder. Coupled-mode theory for optical fibers. *JOSA*, 62(11):1267–1277, Nov. 1972.

[175] Pierre Soille. *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

[176] Geoffrey A Sonn, Sha-Nita E Jones, Tatum V Tarin, Christine B Du, Kathleen E Mach, Kristin C Jensen, and Joseph C Liao. Optical biopsy of human bladder neoplasia with in vivo confocal laser endomicroscopy. *The Journal of urology*, 182(4):1299–1305, Jan 2009.

[177] Filip Sroubek and Jan Flusser. Multichannel blind iterative image restoration. *IEEE Trans. on Image Process.*, 12(9):1094–1106, March 2003.

[178] Filip Sroubek and Jan Flusser. Multichannel blind deconvolution of spatially misaligned images. *IEEE Trans. on Image Process.*, 14(7):874–883, July 2005.

[179] Thomas G Stockham, Thomas M Cannon, and Robert B Ingebretsen. Blind deconvolution through digital signal processing. *Proceedings of the IEEE*, 63(4):678–692, April 1975.

[180] P Su, Y Liu, S Lin, K Xiao, P Chen, S An, J He, and Y Bai. Efficacy of confocal laser endomicroscopy for discriminating colorectal neoplasms from non-neoplasms: a systematic review and meta-analysis. *Colorectal Disease*, 15(1):e1–e12, Jan 2013.

[181] Luc Thiberville, Sophie Moreno-Swirc, Tom Vercauteren, Eric Peltier, Charlotte Cavé, and Genevieve Bourg Heckly. In vivo imaging of the bronchial wall microstructure using fibered confocal fluorescence microscopy. *American journal of respiratory and critical care medicine*, 175(1):22–31, Jan 2007.

[182] Luc Thiberville, Mathieu Salaun, Sammy Lachkar, Stephane Dominique, Sophie Moreno-Swirc, Christine Vever-Bizet, and Genevieve Bourg-Heckly. Human in vivo fluorescence microimaging of the alveolar ducts and sacs during bronchoscopy. *European Respiratory Journal*, 33(5):974–985, 2009.

[183] Luc Thiberville, Mathieu Salaün, Samy Lachkar, Stephane Dominique, Sophie Moreno-Swirc, Christine Vever-Bizet, and Genevieve Bourg-Heckly. Confocal fluorescence endomicroscopy of the human airways. *Proceedings of the American Thoracic Society*, 6(5):444–449, Sept. 2009.

[184] Luc Thiberville, Mathieu Salaün, Samy Lachkar, Stéphane Dominique, Sophie Moreno-Swirc, Christine Vever-Bizet, and Geneviève Bourg-Heckly. In vivo confocal fluorescence endomicroscopy of lung cancer. In *13th world conference on lung cancer*, pages 1–1, San Francisco, United States, Aug. 2009.

[185] Dominik Thomann, Daniel R Rines, Peter K Sorger, and Gaudenez Danuser. Automatic fluorescent tag detection in 3d with super-resolution: application to the analysis of chromosome movement. *Journal of microscopy*, 208(1):49–64, Oct 2002.

[186] Tong Tong, Robin Wolz, Pierrick Coupé, Joseph V Hajnal, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Segmentation of mr images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *NeuroImage*, 76:11–23, Aug 2013.

[187] Joshua A Udovich, Andrew R Rouse, Anthony Tanbakuchi, Molly A Brewer, Richard Sampliner, and Arthur F Gmitro. Confocal microendoscope for use in a clinical setting. In *Endoscopic Microscopy II*, volume 6432, pages 64320H1 – 64320H9. International Society for Optics and Photonics, Feb 2007.

[188] David A Van Dyk and Taeyoung Park. Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103(482):790–796, June 2008.

[189] Tom Vercauteren. *Image registration and mosaicing for dynamic in vivo fibered confocal microscopy*. PhD thesis, École Nationale Supérieure des Mines de Paris, 2008.

[190] Tom Vercauteren, Aymeric Perchant, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical image analysis*, 10(5):673–692, 2006.

[191] Konstantinos Veropoulos, C. Campbell, G. Learmonth, B. Knight, and J. Simpson. The automated identification of tubercle bacilli using image processing and neural computing techniques. In *International Conference on Artificial Neural Networks (ICANN)*, pages 797–802, Skovde, Sweden, Sept 1998. Springer.

[192] Salvador Villena, Miguel Vega, Rafael Molina, and Aggelos K Katsaggelos. A general sparse image prior combination in super-resolution. In *IEEE International Conference on Digital Signal Processing (DSP)*, pages 1–6, Santorini, Greece, 2013.

[193] Matthias von Borstel, Melih Kandemir, Philip Schmidt, Madhavi K Rao, Kumar Rajamani, and Fred A Hamprecht. Gaussian process density counting from weak supervision. In *European Conference on Computer Vision (ECCV)*, pages 365–380, Amsterdam, The Netherlands, Oct 2016. Springer.

[194] Michael B Wallace and Paul Fockens. Probe-based confocal laser endomicroscopy. *Gastroenterology*, 136(5):1509–1513, Sept 2009.

[195] Jiafu Wang, Min Yang, Li Yang, Yun Zhang, Jing Yuan, Qian Liu, Xiaohua Hou, and Ling Fu. A confocal endoscope for cellular imaging. *Engineering*, 1(3):351–360, Sept 2015.

[196] Kenneth K Wang, David L Carr-Locke, Satish K Singh, Helmut Neumann, Helga Bertani, Jean-Paul Galmiche, Razvan I Arsenescu, Fabrice Caillol, Kenneth J Chang, Stanislas Chaussade, et al. Use of probe-based confocal laser endomicroscopy (pcle) in gastrointestinal applications. a consensus report based on clinical evidence. *United European gastroenterology journal*, 3(3):230–254, June 2015.

[197] Ralph Weissleder. *Molecular imaging: principles and practice*. PMPH-USA, 2010.

[198] David Wipf and Haichao Zhang. Revisiting bayesian blind deconvolution. *The Journal of Machine Learning Research*, 15(1):3595–3634, Nov 2014.

[199] Chee Sun Won and Robert M Gray. *Stochastic image processing*. Springer Science & Business Media, 2013.

[200] Alexander Wong, Xiao Yu Wang, and Maud Gorbet. Bayesian-based deconvolution fluorescence microscopy using dynamically updated nonstationary expectation estimates. *Scientific reports*, 5, June 2015.

[201] Harry AC Wood, Kerrianne Harrington, James M Stone, Tim A Birks, and Jonathan C Knight. Quantitative characterization of endoscopic imaging fibers. *Optics express*, 25(3):1985–1992, Feb. 2017.

[202] Katherine Wu, Jen-Jane Liu, Winifred Adams, Geoffrey A Sonn, Kathleen E Mach, Ying Pan, Andrew H Beck, Kristin C Jensen, and Joseph C Liao. Dynamic real-time microscopy of the urinary tract using confocal laser endomicroscopy. *Urology*, 78(1):225–231, May 2011.

[203] Weidi Xie, J Alison Noble, and Andrew Zisserman. Microscopy cell counting and detection with fully convolutional regression networks. *Computer methods in biomechanics and biomedical engineering: Imaging & Visualization*, 6(3):283–292, 2018.

[204] Y Xue and N Ray. Cell detection in microscopy images with deep convolutional neural network and compressed sensing. *Submitted IEEE Transactions on Image Processing*, 2018.

[205] Jianchao Yang, John Wright, Thomas S Huang, Yi Ma, et al. Image super-resolution as sparse representation of raw image patches. In *CVPR*, volume 1, page 3, Anchorage, AK, USA, June 2008.

[206] Yu-Li You and Mostafa Kaveh. Blind image restoration by anisotropic regularization. *IEEE Trans. on Image Process.*, 8(3):396–407, March 1999.

[207] Jonas Yserbyt, Christophe Dooms, Marc Decramer, and Geert M Verleden. Acute lung allograft rejection: Diagnostic role of probe-based confocal laser endomicroscopy of the respiratory tract. *The Journal of Heart and Lung Transplantation*, 33(5):492–498, May 2014.

[208] Jonas Yserbyt, Christophe Dooms, Wim Janssens, and GM Verleden. Endoscopic advanced imaging of the respiratory tract: exploring probe-based confocal laser endomicroscopy in emphysema. *Thorax*, pages thoraxjnl–2016, April 2017.

[209] Jonas Yserbyt, Christophe Dooms, Vincent Ninane, Marc Decramer, and Geert Verleden. Perspectives using probe-based confocal laser endomicroscopy of the respiratory tract. *Swiss Med Wkly*, 143, March 2013.

[210] Chengpu Yu, Cishen Zhang, and Lihua Xie. A blind deconvolution approach to ultrasound imaging. *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, 59(2):271–280, February 2012.

[211] Riccardo Zanella, G Zanghirati, R Cavicchioli, L Zanni, P Boccacci, M Bertero, and G Vicidomini. Towards real-time image deconvolution: application to confocal and sted microscopy. *Scientific reports*, 3:1–8, Aug. 2013.

[212] Aqib H Zehri, Joseph F Georges Wyatt Ramey, Michael A Mooney, Nikolay L Martirosyan, Mark C Preul, and Peter Nakaji. Neurosurgical confocal endomicroscopy: a review of contrast agents, confocal systems, and future imaging modalities. *Surgical neurology international*, 5, April 2014.

[213] Jun Zhang. The mean field theory in em procedures for blind markov random field image restoration. *IEEE Trans. on Image Process.*, 2(1):27–40, Jan 1993.

[214] Ningning Zhao, Adrian Basarab, Denis Kouamé, and Jean-Yves Tourneret. Joint segmentation and deconvolution of ultrasound images using a hierarchical bayesian model based on generalized gaussian priors. *IEEE Trans. on Image Process.*, 25(8):3736–3750, Aug 2016.